Joint String Complexity for Markov Sources: Small Data Matters *

Philippe Jacquet, Dimitris Milioris, Wojciech Szpankowski

Joint String Complexity for Markov Sources: Small Data Matters*

Philippe Jacquet
INRIA
Paris
France
Philippe.Jacquet@inria.fr

Dimitris Milioris
Bell Labs - Nokia
91620 Nozay
France
dimitrios.milioris@nokia-bell-labs.com

Wojciech Szpankowski
Dept. of Computer Sciences
Purdue University
West Lafayette, IN 47907
spa@cs.purdue.edu

Abstract

String complexity is defined as the cardinality of a set of all distinct words (factors) of a given string. For two strings, we introduce the *joint string complexity* as the cardinality of a set of words that are common to both strings. String complexity finds a number of applications from capturing the richness of a language to finding similarities between two genome sequences. In this paper we analyze the joint string complexity when both strings are generated by Markov sources. We prove that the joint string complexity grows linearly (in terms of the string lengths) when both sources are statistically indistinguishable and sublinearly when sources are statistically distinguishable. Precise analysis of the joint string complexity requires subtle singularity analysis and saddle point method over infinity many saddle points leading to novel oscillatory phenomena with single and double periodicities. To overcome these challenges, we apply analytic techniques such as multivariate generating functions, multivariate depoissonization and Mellin transform, spectral matrix analysis, and complex asymptotic methods.

Index terms: String complexity, joint string complexity, suffix trees, Markov sources, source discrimination, generating functions, Mellin transform, saddle point methods, analytic information theory.

1 Introduction

In the last decades, several attempts have been made to capture mathematically the concept of "complexity" of a sequence. The notion is connected with quite deep mathematical properties, including rather elusive concept of randomness in a string (see e.g., [5, 16, 19]), and the "richness of the language". The *string complexity* is defined as the number of *distinct substrings* of the underlying string. More precisely, if X is a sequence and $I(X)$ is its set of factors (distinct subwords), then the cardinality $|I(X)|$ is the complexity of the sequence. For example, if $X = aabaa$ then $I(X) = \{\nu, a, b, aa, ab, ba, aab, aba, baa, aaba, abaa, aabaa\}$ and $|I(X)| = 12$ (ν denotes the empty string). Sometimes the complexity of a string is called the I-complexity [3]. This measure is simple but quite intuitive. Sequences with low complexity contain a large number of repeated substrings and they eventually become periodic.

In general, however, information contained in a string cannot be measured in absolute and a reference string is required. To this end we introduced in [6] the concept of the *joint* string complexity, or J-complexity, of two strings. The J-complexity is the number of *common distinct factors* in two sequences. In other words, the J-complexity of sequences X and Y is equal to $J(X, Y) = |I(X) \cap I(Y)|$. We denote by $J_{n,m}$ the *average* value of $J(X, Y)$ when X is of length n and Y is of length m. In this paper, we study the joint string complexity for Markov sources when $n = m$.

The J-complexity is an efficient way of estimating similarity degree of two strings. For example, genome sequences of two dogs will contain more common words than genome sequences of a dog and a

[0]This work was partially supported by LINCS. Paris as well as the NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grants CCF-1524312, CCF-2006440, and CCF-2007238.

cat. Similarly, two texts written in the same language have more words in common than texts written in very different languages. Thus, the J-complexity is larger when languages are close (*e.g.* French and Italian), and smaller when languages are different (*e.g.* English and Polish); see Figures 1-2. In fact, texts in the same language but on different topics (*e.g.* law and cooking) have smaller J-complexity than texts on the same topic (*e.g.* medicine). Furthermore, string complexity has a variety of applications in detection of similarity degree of two sequences, for example "copy-paste" in texts or documents that will allow to detect plagiarism. It could also be used in analysis of social networks (e.g. tweets that are limited to 140 characters) and classification. Therefore it could be a pertinent tool for automated monitoring of social networks (see [1, 18] for some experimental results). However, real time search in blogs, tweets and other social media must balance quality and relevance of the content, which – due to short but frequent posts – is still an unsolved problem. For these short texts precise analysis is highly desirable. We call it the "small data" problem and we hope our rigorous asymptotic analysis of the joint string complexity will shed some light on this problem. In this paper we offer a precise analysis of the joint complexity together with some experimental results (cf. Figures 1 and 2) confirming usefulness of the joint string complexity for text discrimination. To better model real texts, we assume that both sequences are generated by Markov sources making the analysis quite involved. To overcome these difficulties we shall use powerful analytic techniques such as multivariate generating functions, multivariate depoissonization and Mellin transform, spectral matrix analysis, and saddle point methods.

String complexity was studied extensively in the past. The literature is reviewed in [13] where precise analysis of string complexity is discussed for strings generated by unbiased memoryless sources. Another analysis of the same situation was also proposed in [6] where for the first time the joint string complexity for memoryless sources was presented. It was evident from [6] that precise analysis of the joint complexity is quite subtle requiring singularity analysis and infinite number of saddle points. In this paper we deal with the joint string complexity for Markov sources. To the best of our knowledge this problem was never tackled before except in our recent conference paper [9]. As expected, its analysis is rather sophisticated but at the same time quite rewarding. It requires generalized (two-dimensional) dePoissonization and generalized (two-dimensional) Mellin transforms.

In [6] it is proved that the J-complexity of two texts generated by two *different* binary memoryless sources grows as

$$\gamma\frac{n^\kappa}{\sqrt{\alpha\log n}}$$

for some $\kappa < 1$ and $\gamma, \alpha > 0$ depending on the parameters of the sources. When sources are statistically identical, then the J-complexity growth is $O(n)$, hence $\kappa = 1$. When the texts are identical (i.e, $X = Y$), then the J-complexity is identical to the I-complexity and it grows as $\frac{n^2}{2}$ [13]. Indeed, the presence of a common factor of length $O(n)$ inflates the J-complexity to $O(n^2)$, where n is the string length.

We should point out that our experiments indicate a very slow convergence of the complexity estimates for memoryless sources. Furthermore, memoryless sources are not appropriate for modeling many sources, e.g., natural languages. In this paper and [9] we extend the J-complexity estimates to Markov sources of any order for a finite alphabet. Although Markov models are no more realistic in some applications than memoryless sources, they seem to be fairly good approximation for text generation.

Here, we derive a second order asymptotics for J-complexity for Markov sources of the following form

$$\gamma\frac{n^\kappa}{\sqrt{\alpha\log n+\beta}}$$

for some $\beta, \gamma > 0$ with $\kappa \leq 1$. This new estimate converges faster, although for small text lengths of order $n \approx 10^2$ one needs to compute additional terms of the asymptotic expansion. In fact, for some Markov sources our analysis indicates that J-complexity oscillates with n. This is manifested by the appearance of a periodic function in the leading term of our asymptotics. Surprisingly, this additional term even further improves the convergence for small values of n.

Let us now summarize in full our main results Theorems 1–9. In our first main result Theorem 1 we observe that the joint string complexity $J_{n,m}$ can be asymptotically analyzed by considering a simpler quantity called the *joint prefix complexity* denoted as $C_{n,m}$. To define it, let $\mathcal{X}$ be a set of infinite strings, and we define the *prefix set* $\mathcal{I}(\mathcal{X})$ of $\mathcal{X}$ as the set of prefixes of $\mathcal{X}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be now two sets of strings and we define the *joint prefix complexity* as the number of common prefixes, *i.e.* $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{Y})|$. When

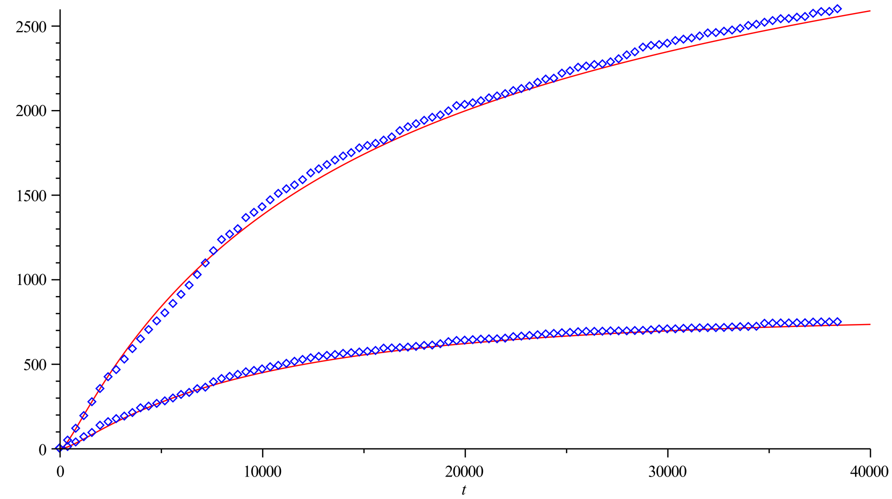

Figure 1: Joint complexity of simulated texts (3rd Markov order) of English, vs French (top), Polish (bottom) languages, versus average theoretical (plain). The x-axis is the length of the text, while the y-axis shows the joint complexity.

$\mathcal{X}_n$ is a set of n independently generated strings by Markov source 1 and $\mathcal{Y}_m$ is a set of m independently generated strings by source 2, then $C_{n,m}$ is $C_{n,m} = \mathbf{E}|\mathcal{I}(\mathcal{X}_n) \cap \mathcal{I}(\mathcal{Y}_m)|) - 1$ which represents the number of common prefixes between $\mathcal{X}_n$ and $\mathcal{Y}_m$. In the remaining part of the paper we only deal with the joint prefix complexity $C_{n,m}$. We show in Theorem 1 that $|J_{n,m} - C_{n,m}| = O(n^\epsilon + m^\epsilon)$ for some $\epsilon > 0$, suggesting that may focus our attention on simpler to analyze quantity, namely $C_{n,m}$.

First, in Theorem 2 we considered two statistically identical sources and prove that the joint string complexity grows linearly with n: For certain sources called *noncommensurable* there is a constant in front of n (that we determine) while for *commensurable sources* the factor in front of n is a *fluctuating periodic function* of small amplitude. We shall see these two cases permeate all our results. Then we deal in Theorem 3 with a special sources in which the underlying Markov matrices are nilpotent. After that we study general sources, however, we split our presentation and proofs into two parts. First, in Theorems 4–5 we assume that one of the source is uniform. Under this assumption we develop techniques to prove our results. Finally, in Theorem $7-9$ we study the general case.

Let us now compare our theoretical results with experimental results on real/simulated texts generated in different languages. In Figure 1 we compare the joint complexity of a simulated English text with the same length texts simulated in French and in Polish. In the simulation we use a Markov model of order 3. It is easy to see that even for texts of lengths smaller than a thousand one can discriminate between these languages. In fact, computations show that for English versus French we have $\kappa = 0.18$; and versus Polish: $\kappa = 0.1$, Furthermore, for a Markov model of order 3 we find that English text has entropy (per symbol): 0.944; French: 0.934; Polish: 0.665. The theoretical curves shown in Figure 1 are obtained through Theorem 8, however, for small values of the text length it is computed via the iterative resolution of functional equations (49) and (51). Figure 2 shows the continuation of our theoretical estimates up to $n = 10^{10}$ and compared with the theoretical estimate $O(n^\kappa)$ as presented in Theorems $7-8$.

The joint string complexity can be used to discriminate Markov sources [24] since, as already observed, the growth of the joint string complexity is $O(n^\kappa)$ with $\kappa = 1$ when sources are statistically indistinguishable and $\kappa < 1$ otherwise. For example, we can use the joint string complexity to verify authorship of an unknown manuscript by comparing it to a manuscript of known authorship and checking whether $\kappa = 1$ or not. More precisely, we propose to introduce the following discriminant function

$$d(X,Y) = 1 - \frac{1}{\log n} \log J(X,Y)$$

for two sequences X and Y of length n. This discriminant allows us to determine whether X and Y

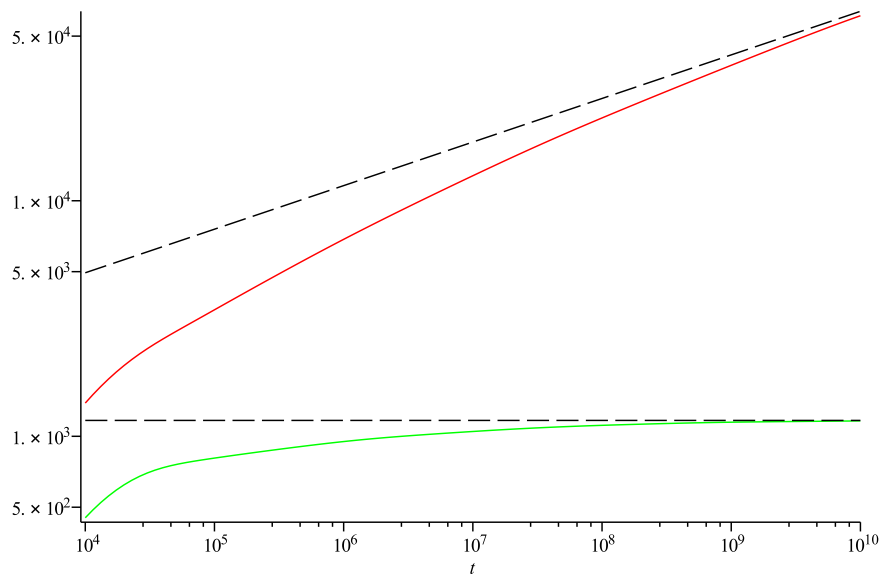

Figure 2: Average theoretical joint complexity of 3rd Markov order text of English, vs French (top), and vs Polish (bottom) languages, versus order estimate $O(n^\kappa)$. The x-axis is the length of the text, while the y-axis shows the joint complexity.

are generated by the same Markov source or not by verifying whether $d(X,Y) = O(1/\log n) \to 0$ or $d(X,Y) = 1 - \kappa + O(\log\log n/\log n) > 0$, respectively. In fact, we used it with some success to classify tweets (see SNOW 2014 challenge of tweets classification and topic detection [1]).

The paper is organized as follows. In the next section we present our main results Theorems 1–9. We prove Theorem 1 in Section 3. Then we present some preliminary results in Section 4. In particular, we derive the functional equation for the joint prefix complexity $C_{n,m}$, establish some depoissonization results, and derive double Mellin transform. We first prove the nilpotent case in Section 5. The proofs of Theorems 4 – 5 are presented in Section 6, and the proofs of Theorem 7 – 9 are discussed in Section 7.

2 Main Results

In this section we define precisely our problem, introduce some important notation, and present our main results Theorem 1 – Theorem 9. The proofs are presented in the remaining parts of the paper and Appendix.

2.1 Models and notations

We begin by introducing some general notation. Let X and w be two strings over the alphabet $\mathcal{A}$. We denote by $|X|_w$ the number of times w occurs in X (e.g., $|abbba|_{bb} = 2$ when $X = abbba$ and $w = bb$). By convention $|X|_\nu = |X| + 1$, where ν is the empty string.

Throughout we denote by X a string (text) whose complexity we plan to study. We also assume that its length $|X|$ is equal to n. Then we define $I(X) = \{w : \ |X|_w \geq 1\}$, that is, $I(X)$ contains all *distinct* subwords of X. Observe that the string complexity $|I(X)|$ can be represented as

$$|I(X)| = \sum_{w \in \mathcal{A}^*} 1_{|X|_w \geq 1},$$

where 1_A is the indicator function of the event A. Notice that $|I(X)|$ is equal to the number of nodes in the associated suffix tree of X [13, 23] (see also [7]).

Now, let X and Y be two strings (not necessarily of the same length). We define the *joint string complexity* as the cardinality of the set $J(X,Y) = I(X) \cap I(Y)$, that is,

$$|J(X,Y)| = \sum_{w \in \mathcal{A}^*} 1_{|X|_w \geq 1} \times 1_{|Y|_w \geq 1} \ .$$

In other words, $J(X,Y)$ represents the number of *common* and distinct subwords of both X and Y. For example, if $X = aabaa$ and $Y = abbba$, then $J(X,Y) = \{\nu, a, b, ab, ba\}$.

In this paper we assume that both strings X and Y are generated by two *independent Markov sources* of order r (we will only deal here with Markov of order 1, but extension to arbitrary order is straightforward). We assume that source i, for $i \in \{1,2\}$ has the transition probabilities $P_i(a|b)$ from state b to state a, where $a, b \in \mathcal{A}^r$. We denote by $\mathbf{P}_1$ (resp. $\mathbf{P}_2$) the transition matrix of Markov source 1 (resp. source 2). The stationary distributions are respectively denoted by $\boldsymbol{\pi}_1(a)$ and $\boldsymbol{\pi}_2(a)$ for $a \in \mathcal{A}^r$. Throughout, we consider general Markov sources with transition matrices $\mathbf{P}_i$ that may contain zero coefficients. This assumption leads to interesting embellishment of our results.

Let X_n and Y_m be two strings of respective lengths n and m, generated by Markov source 1 and Markov source 2, respectively. We write

$$J_{n,m} = \mathbf{E}(|J(X_n, Y_m)|) - 1 = \sum_{w \in \mathcal{A}^* \setminus \nu} P(|X|_w \geq 1) P(|Y|_w \geq 1) \tag{1}$$

for the joint complexity, *i.e.* omitting the empty string. In this paper we study $J_{n,m}$ for $n = \Theta(m)$.

It turns out that analyzing $J_{n,m}$ is very challenging. It is related to the number of common nodes in two suffix trees, one built for X and the other built for Y. We know that analysis of a single suffix tree is quite challenging [7, 20]. Its analysis is reduced to study a simpler structure known as *tries*, a digital tree built from prefixes of a set of *independent* strings. We shall follow this approach here. Therefore, we introduce another concept. Let $\mathcal{X}$ be a set of infinite strings, and we define the prefix set $\mathcal{I}(\mathcal{X})$ of $\mathcal{X}$ as the set of prefixes of $\mathcal{X}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be now two sets of strings and we define the *joint prefix complexity* as the number of common prefixes, *i.e.* $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{Y})|$. When $\mathcal{X}_n$ is a set of n independent strings generated by source 1 and $\mathcal{Y}_m$ is a set of m independent strings generated by source 2, then we define $C_{n,m}$ as

$$C_{n,m} = \mathbf{E}|\mathcal{I}(\mathcal{X}_n) \cap \mathcal{I}(\mathcal{Y}_m)|) - 1$$

which represents the number of common prefixes between $\mathcal{X}_n$ and $\mathcal{Y}_m$.

Observe that we can re-write $C_{n,m}$ in a different way. Define $\Omega^i(w)$ for $i = 1,2$ as the number of strings in $\mathcal{X}$ and $\mathcal{Y}$, respectively, whose prefixes are equal to w provided that strings in $\mathcal{X}$ are generated by source 1 and strings in $\mathcal{Y}$ by source 2. Then, it is easy to notice that

$$C_{n,m} = \sum_{w \in \mathcal{A}^* \setminus \nu} P(\Omega^1_n(w) \geq 1) P(\Omega^2_m(w) \geq 1) \tag{2}$$

which should be compared to (1).

The idea is that $C_{n,m}$ is a good approximation of $J_{n,m}$ as we present in our first main result Theorem 1. We shall see in Sections 4 – 7 that $C_{n,m}$ are easier to analyze, however, far from simple. In fact, $C_{n,m}$ has a nice interpretation. It corresponds to the number of common nodes in two tries built from $\mathcal{X}$ and $\mathcal{Y}$. We know [11, 10, 23] that tries are easier to analyze than suffix trees.

2.2 Summary of Main Results

We now present our main theoretical results. In the first foundational result below we show that asymptotically we can analyze $J_{n,m}$ through the quantity $C_{n,m}$ defined above in (2). The proof can be found in Section 3.

Theorem 1. *Let n and m be of the same order. Then there exists $1/2 \leq \epsilon < 1$ such that*

$$J_{n,m} = C_{n,m} + O(n^\epsilon + m^\epsilon) \tag{3}$$

as $n \to \infty$.

In the rest of the paper we shall analyze $C_{n,n}$. We should point out that the error term could be as large as the leading term, but for sources that are relatively close the error term will be negligible.

Now we present a series of results each treating different cases of Markov sources. Moreover, our results depend on whether the underlying Markov sources are commensurable or not so we next define them.

Definition 1 (Rationally Related Matrix). *We say that a matrix* $\mathbf{M} = [m_{ab}]_{(a,b)\in\mathcal{A}^2}$ *is rationally related if* $\forall(a,b,c) \in \mathcal{A}^3$ *we have* $m_{ab} + m_{ca} - m_{cb} \in \mathbb{Z}$*, where* $\mathbb{Z}$ *is the set of integers.*

Definition 2 (Logarithmically rationally related matrix). *We say that a matrix* $\mathbf{M} = [m_{ab}]$ *is logarithmically rationally related if there exists a non zero real number* x *such that the matrix* $x\log^*(\mathbf{M})$ *is rationally related, where the matrix* $\log^*(\mathbf{M})$ *is composed of* $\log(m_{ab})$ *when* $m_{ab} > 0$ *and zero otherwise. The smallest non negative value* ω *of the real* x *defined above is called the root of* $\mathbf{M}$*.*

The following matrix is an example of logarithmically rationally related matrix:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{4} & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}.$$

Its root is $1/\log 2$.

Definition 3 (Logarithmically commensurable pair). *We say that a pair of two matrices* $\mathbf{M} = [m_{ab}]_{(a,b)\in\mathcal{A}^2}$ *and* $\mathbf{M}' = [m'_{ab}]$ *is logarithmically commensurable if there exists a pair of real numbers* (x, y) *such that* $x\log^*(\mathbf{M}) + y\log^*(\mathbf{M}')$ *is not null and is logarithmically rationally related.*

Notice that when $\mathbf{M}$ and $\mathbf{M}'$ are both rationally related, then the pair is logarithmically commensurable. Nevertheless it is possible to have logarithmically commensurable pairs with the individual matrices not logarithmically rationally related. For example when $\log^*\mathbf{M}' = 2\pi\mathbf{Q} + \log^*\mathbf{M}$ with $\mathbf{Q}$ an integer matrix.

We are now in the position to discuss our first main result for Markov sources that are statistically indistinguishable. Throughout we present results for $m = n$.

Theorem 2. *Consider the average joint complexity of two texts of length* n *generated by the same general stationary Markov source, that is,* $\mathbf{P} := \mathbf{P}_1 = \mathbf{P}_2$*.*
(i) *[Noncommensurable Case.] Assume that* $\mathbf{P}$ *is not logarithmically rationally related. Then*

$$J_{n,n} = \frac{2n\log 2}{h} + o(n) \tag{4}$$

where h *is the entropy rate of the source defined as* $h = \sum_{a,b\in\mathcal{A}} \pi(a)P(a|b)$*.*
(ii) *[Commensurable Case.] Assume that* $\mathbf{P}$ *is logarithmically rationally related. Then there is* $\epsilon < 1$ *such that:*

$$J_{n,n} = \frac{2n\log 2}{h}(1 + Q_0(\log n)) + O(n^\epsilon) \tag{5}$$

where $Q_0(.)$ *is a periodic function of small amplitude. (In Section 4.3 we compute explicitly* Q*.)*

Now we consider sources that are not the same and have respective transition matrices $\mathbf{P}_1$ and $\mathbf{P}_2$. The transition matrices are on $\mathcal{A}^r \times \mathcal{A}^r$, however, hereafter we deal mostly with $r = 1$. If $(a, b) \in \mathcal{A}\times\mathcal{A}$, we denote by $P_i(a|b)$ the (a, b)-th coefficient of matrix $\mathbf{P}_i$. For a tuple of complex numbers (s_1, s_2) we write $\mathbf{P}(s_1, s_2)$ for the following matrix

$$\mathbf{P}(s_1, s_2) = [(\mathbf{P}_1(a|b))^{-s_1}(\mathbf{P}_2(a|b))^{-s_2}]_{a,b\in A}.$$

In fact, we can write it as the Schur product, denoted as $\star$, of two matrices $\mathbf{P}_1(s_1) = (\mathbf{P}_1(a|b))^{-s_1}$ and $\mathbf{P}_2(s_2) = (\mathbf{P}_2(a|b))^{-s_2}$, that is, $\mathbf{P}(s_1, s_2) = \mathbf{P}_1(s_1) \star \mathbf{P}_2(s_2)$.

To present succinctly our general results we need some more notation. Let $\langle\mathbf{x}|\mathbf{y}\rangle$ be the scalar product of vector $\mathbf{x}$ and vector $\mathbf{y}$. By $\lambda(s_1, s_2)$ we denote the main eigenvalue of matrix $\mathbf{P}(s_1, s_2)$, and

$\mathbf{u}(s_1, s_2)$ its corresponding right eigenvector (i.e, $\lambda(s_1, s_2)\mathbf{u}(s_1, s_2) = \mathbf{P}(s_1, s_2)\mathbf{u}(s_1, s_2)$), and $\boldsymbol{\zeta}(s_1, s_2)$ its left eigenvector (i.e., $\lambda(s_1, s_2)\boldsymbol{\zeta}(s_1, s_2) = \boldsymbol{\zeta}(s_1, s_2)\mathbf{P}(s_1, s_2)$). We assume that $\langle \boldsymbol{\zeta}(s_1, s_2) | \mathbf{u}(s_1, s_2) \rangle = 1$. Furthermore, the vector $\boldsymbol{\pi}(s_1, s_2)$ is defined as the vector $(\pi_1(a)^{-s_1}\pi_2(a)^{-s_2})_{a\in\mathcal{A}}$ where $(\pi_i(a))_{a\in\mathcal{A}}$ is the left eigenvector of matrix $\mathbf{P}_i$ for $i \in \{1, 2\}$. In other words $(\pi_i(a))_{a\in\mathcal{A}}$ is the stationary distribution of the Markov source i.

We start our presentation with the simplest case, namely the case when the matrix $\mathbf{P}(0, 0)$ is nilpotent [14], that is, for some K the matrix $\mathbf{P}^K(0, 0)$ is the null matrix. Notice that for nilpotent matrices $\forall(s_1, s_2)$: $\mathbf{P}^K(s_1, s_2) = 0$.

Theorem 3. *If $\mathbf{P}(s_1, s_2)$ is nilpotent, then there exists γ_0 such that*

$$\lim_{n\to\infty} J_{n,n} = \gamma_0 := \langle \mathbf{1}_C (\mathbf{I} - \mathbf{P}(0, 0))^{-1} | \mathbf{1} \rangle \tag{6}$$

where $\mathbf{1}$ is the unit vector, $\mathbf{1}_C$ the vector on $\mathcal{A}$ with $\mathbf{1}_C(a) = 1$ when a is common to both sources, and $\mathbf{1}_C(a) = 0$ otherwise.

This result is not surprising and rather trivial since the common factors can only occur in a finite window at the beginning of the strings. It turns out that $\gamma_0 = 1168$ for 3rd order Markov model of English versus Polish languages used in our experiments.

From now on, we assume that $\mathbf{P}(s_1, s_2)$ is not nilpotent. We need to pay much closer attention to the structure of the set of roots of the *characteristic equation*

$$\lambda(s_1, s_2) = 1$$

that will play a major role in the analysis. We discuss in depth properties of these roots in Section 4.5. Here we introduce only a few important definitions.

Definition 4. *The kernel $\overline{\mathcal{K}}$ is the set of complex tuples (s_1, s_2) such that $\mathbf{P}(s_1, s_2)$ has its largest eigenvalue equal to 1. The real kernel $\mathcal{K} = \overline{\mathcal{K}} \cap \mathbb{R}^2$, i.e. the set of real tuples (s_1, s_2) such that the main eigenvalue $\lambda(s_1, s_2) = 1$.*

The following lemma is easy to prove.

Lemma 1. *The real kernel forms a concave curve in $\mathbb{R}^2$.*

Furthermore, we introduce two important notations:

$$\begin{aligned} \kappa &= \min_{(s_1,s_2)\in\mathcal{K}} \{-s_1 - s_2\} & (7)\\ (c_1, c_2) &= \arg \min_{(s_1,s_2)\in\mathcal{K}} \{-s_1 - s_2\}. & (8) \end{aligned}$$

Easy algebra shows that $\kappa \leq 1$. Furthermore, in the Appendix we prove the following property.

Lemma 2. *Let c_1 and c_2 minimize $-s_1 - s_2$ where real tuple $(s_1, s_2) \in \mathcal{K}$. Assume $\forall(a, b) \in \mathcal{A}^2$: $P_1(a|b) > 0$ then $c_1 \leq 0$ and $c_2 \geq -1$.*

Finally, we introduce a new concept $\partial\mathcal{K}$, the border of the kernel $\overline{\mathcal{K}}$, defined as follows.

Definition 5. *We denote $\partial\mathcal{K}$ the subset of $\overline{\mathcal{K}}$ made of the pairs (s_1, s_2) such $\Re(s_1, s_2) = (c_1, c_2)$.*

The case when both matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ have some zero coefficients is the most intricate part. Therefore, to present our strongest results, we start with a special case when one of the source is uniform. Later we generalize it.

We first consider a special case when source 1 is uniform memoryless, *i.e.* $\mathbf{P}_1 = \frac{1}{|\mathcal{A}|}\mathbf{1} \otimes \mathbf{1}$ and the other matrix $\mathbf{P}_2$ is not nilpotent and general (that is, it may have some zero coefficients). In this case we always have $c_1 < 0$ and $c_2 < 0$. This case we have the following theorem.

Theorem 4. *Let* $\mathbf{P}_1 = \frac{1}{|\mathcal{A}|}\mathbf{1}\otimes\mathbf{1}$ *and* $\mathbf{P}_2 \neq \mathbf{P}_1$ *is a general transition matrix. Thus both* c_1 *and* c_2 *are between* -1 *and* 0.
(i) *[Mono periodic case.] If* $\mathbf{P}_2$ *is not logarithmically rationally related, then there exists a periodic function* $Q_1(x)$ *of small amplitude such that*

$$C_{n,n} = \frac{\gamma_2 n^\kappa}{\sqrt{\alpha_2 \log n + \beta_2}}(1 + Q_1(\log n) + o(1)). \tag{9}$$

(ii) *[Double periodic case.] If* $\mathbf{P}_2$ *is logarithmically rationally related, then there exists a double periodic function*[1] $Q_2(.)$ *of small amplitude such that*

$$C_{n,n} = \frac{\gamma_2 n^\kappa}{\sqrt{\alpha_2 \log n + \beta_2}}(1 + Q_2(\log n) + o(1)). \tag{10}$$

The constants γ_2, α_2 and β_2 in the Theorem 4 are explicitly computable as presented next. To simplify our notation for all $(a,b)\in\mathcal{A}^2$ we shall write $P_2(a|b) = P(a|b)$ and $\mathbf{P}_2 = \mathbf{P}$. Therefore

$$\mathbf{P}(s_1,s_2) = |\mathcal{A}|^{s_1}\mathbf{P}(s) \tag{11}$$

with $\mathbf{P}(s) = \mathbf{P}(0,s)$. We also write $\pi(a) = \pi_2(a)$ and $\boldsymbol{\pi}(s) = [\pi(a)^{-s}]_{a\in\mathcal{A}}$, thus

$$\boldsymbol{\pi}(s_1,s_2) = |\mathcal{A}|^{s_1}\boldsymbol{\pi}(s_2)\ . \tag{12}$$

Let $\lambda(s_1,s_2)$ be again the main (largest) eigenvalue of $\mathbf{P}(s_1,s_2)$. We have

$$\lambda(s_1,s_2) = |\mathcal{A}|^{s_1}\lambda(s_2) \tag{13}$$

where $\lambda(s)$ is the main eigenvalue of matrix $\mathbf{P}(s)$. We also define $\mathbf{u}(s)$ as the right eigenvector of $\mathbf{P}(s)$ and $\boldsymbol{\zeta}(s)$ as the left eigenvector provided $\langle\boldsymbol{\zeta}(s)|\mathbf{u}(s)\rangle = 1$. It is easy to see that

$$\lambda(s) = \langle\boldsymbol{\zeta}(s)|\mathbf{P}(s)\mathbf{u}(s)\rangle.$$

Now we can express c_1 and c_2 defined in (8) in another way. Notice that if $\lambda(s_1,s_2) = 1$, then in this case

$$s_1 = -\log_{|\mathcal{A}|}\lambda(s_2).$$

Define $L(s) = \log_{|\mathcal{A}|}\lambda(s)$. Then c_2 is the value that minimizes $L(s) - s$, that is,

$$\frac{\lambda'(c_2)}{\lambda(c_2)} = \log|\mathcal{A}|. \tag{14}$$

Also $c_1 = -L(c_2)$ and $\kappa = -c_1 - c_2 = \min_s\{L(s) - s\}$. We have $\kappa \leq 1$ since $L(0) = 1$.
We now can present explicit expression for the constants in Theorem 4.

Theorem 5. *We consider the case* $\mathbf{P}_1 = \frac{1}{|\mathcal{A}|}\mathbf{1}\otimes\mathbf{1}$ *and* $\mathbf{P}_2 \neq \mathbf{P}_1$ *has all non negative coefficients. Let* $f(s) = \langle\boldsymbol{\pi}(s)|\mathbf{u}(s)\rangle$ *and* $g(s) = \langle\boldsymbol{\zeta}(s)|\mathbf{1}\rangle$. *Furthermore, with* $\Psi(s)$ *being the Euler psi function, define* $\alpha_2 = L''(c_2)$ *where* $L(s) = \log_{|\mathcal{A}|}\lambda(s)$, *and*

$$\beta_2(s_1,s_2) = -\alpha_2\left(\Psi(s_1) + \frac{1}{1+s_1} + \log|\mathcal{A}|\right) + \Psi'(s_1) - \frac{1}{(s_1+1)^2} + \Psi'(s_2) - \frac{1}{(s_2+1)^2}$$
$$+ \frac{f''(s_2)}{f(s_2)} - \left(\frac{f'(s_2)}{f(s_2)}\right)^2 + \frac{g''(s_2)}{g(s_2)} - \left(\frac{g'(s_2)}{g(s_2)}\right)^2$$

as well as

$$\gamma(s_1,s_2) = \frac{f(s_1)g(s_2)(s_1+1)\Gamma(s_1)(s_2+1)\Gamma(s_2)}{\lambda(s_2)\log|\mathcal{A}|\sqrt{2\pi}}.$$

[1] We recall that a double periodic function is a function on real numbers that is a sum of two periodic functions of non commensurable periods.

We have $c_1 = -\log_{|\mathcal{A}|} \lambda(c_2)$, *and then*

$$C_{n,n} = n^{\kappa} \frac{\gamma(c_1, c_2)}{\sqrt{\alpha_2 \log n + \beta_2(c_1, c_2)}} + n^{\kappa} Q(\log n) + o(\frac{n^{\kappa}}{\sqrt{\log n}}). \tag{15}$$

The function $Q(x)$ *can be expressed as*

$$Q(x) = \sum_{(s_1,s_2)\in \partial\mathcal{K}^*} e^{ix\Im(s_1+s_2)} \frac{\gamma(s_1, s_2)}{\sqrt{\alpha_2 x + \beta_2(s_1, s_2)}}.$$

If the matrix $\mathbf{P}_2$ *is logarithmically rationally related, then* $\partial\mathcal{K}$ *is a lattice. Let* ω *the root of* $\mathbf{P}_2$ *then*

$$\partial\mathcal{K} = \left\{ \left(c_1 + \frac{2ik\pi}{\log|\mathcal{A}|}, c_2 + 2i\pi\ell\omega \right), (k, \ell) \in \mathbb{Z}^2 \right\},$$

and $\sqrt{x}Q(x)$ *is asymptotically double periodic. Otherwise (i.e., irrational case),*

$$\partial\mathcal{K} = \left\{ \left(c_1 + \frac{2ik\pi}{\log 2}, c_2 \right),\ k \in \mathbb{Z} \right\}$$

and $\sqrt{x}Q(x)$ *is asymptotically single periodic. The amplitude of* Q *is of order* 10^{-6}.

Now we consider the case when the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are general and $\mathbf{P}_1 \neq \mathbf{P}_2$. If they contain some zero coefficients, then we may have $\mathbf{P}(-1, 0) \neq \mathbf{P}_1$ and/or $\mathbf{P}(0, -1) \neq \mathbf{P}_2$. In this (very unlikely) case we may have $\mathbf{P}(-1, 0) = \mathbf{P}(0, -1)$ or more generally they are conjugate.[2] For example for matrices:

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 0 & \frac{5}{8} & \frac{1}{2} \\ 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{2} & \frac{5}{8} & 0 & \frac{1}{8} \end{bmatrix} \mathbf{P}_2 = \begin{bmatrix} \frac{1}{2} & \frac{5}{8} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & \frac{5}{8} & \frac{1}{8} \end{bmatrix}$$

we have

$$\mathbf{P}(-1, 0) = \mathbf{P}(0, -1) = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \\ 0 & 0 & 0 & \frac{1}{8} \end{bmatrix}$$

If $\mathbf{P}(-1, 0) = \mathbf{P}(0, -1)$, then there is no unique solution (c_1, c_2) of the characteristic equation. As a consequence, we have the following theorem:

Theorem 6. *When* $\mathbf{P}(-1, 0)$ *and* $\mathbf{P}(0, -1)$ *are conjugate matrices we have*

$$C_{n,n} = \gamma_0(-\kappa) n^{\kappa} (1 + o(1)) \tag{16}$$

where $\kappa < 1$ *is such that* $(-\kappa, 0) \in \mathcal{K}$ *and* $\gamma_0(-\kappa)$ *are explicitly computable. When both matrices are logarithmically rationally related, then*

$$C_{n,n} = \gamma(-\kappa) n^{\kappa} \left(1 + Q_0(\log n) + O(n^{-\epsilon} \right) \tag{17}$$

where $Q_0(.)$ *is small periodic function and* $\epsilon > 0$.

In general, however, $\mathbf{P}(-1, 0)$ and $\mathbf{P}(0, -1)$ are *not* conjugate, and therefore, there is a unique (c_1, c_2) of the characteristic equation $\lambda(c_1, c_2) = 1$. As in the special case discussed above, $c_1 > -1$ and $c_2 > -1$, however, our results are quantitatively different when $c_1 > 0$ or $c_2 > 0$. We consider it first in Theorem 7 below. Since both cases cannot occur simultaneously, we dwell only on the case $c_2 > 0$; the case $c_1 > 0$ can be handled in a similar manner.

[2] A conjugate matrix is a matrix obtained from a given matrix by taking the complex conjugate of each element of it.

Theorem 7. *Assume* $\mathbf{P}(0,0)$ *is not nilpotent and* $c_2 > 0$.
(i) [Noncommensurable Case.] *We assume that* $\mathbf{P}(-1,0)$ *is not logarithmically related. Let* $-1 < c_0 < 0$ *such that* $(c_0, 0) \in \mathcal{K}$. *There exist* γ_1 *such that*

$$C_{n,n} = \gamma_1 n^{-c_0}(1 + o(1)) \tag{18}$$

(ii) [Commensurable Case.] *Let now* $\mathbf{P}(-1,0)$ *be logarithmically rationally related. There exists a periodic function* $Q_1(.)$ *of small amplitude such that*

$$C_{n,n} = \gamma_1 n^{-c_0}(1 + Q_1(\log n) + O(n^{-\epsilon}) \tag{19}$$

Finally, we handle the most intricate case when both c_1 and c_2 are between -1 and 0. Recall that when both matrix $\mathbf{P}_1$ and $\mathbf{P}_2$ have all positive coefficients, then $\mathbf{P}(-1,0) = \mathbf{P}_1$ and $\mathbf{P}(0,-1) = \mathbf{P}_2$.

Theorem 8. *Assume that both* c_1 *and* c_2 *are between* -1 *and* 0 *and* $\mathbf{P}(0,0)$ *is not nilpotent.*
(i) [Non periodic Case.] *If* $\mathbf{P}(-1,0)$ *and* $\mathbf{P}(0,-1)$ *are not logarithmically commensurable matrices, then there exist* α_2, β_2 *and* γ_2 *such that*

$$C_{n,n} = \frac{\gamma_2 n^{\kappa}}{\sqrt{\alpha_2 \log n + \beta_2}}(1 + o(1)). \tag{20}$$

(ii) [Mono periodic case.] *If only one of the matrices* $\mathbf{P}(-1,0)$ *and* $\mathbf{P}(0,-1)$ *is logarithmically rationally related, then there exists a periodic function* $Q_2(x)$ *of small amplitude such that*

$$C_{n,n} = \frac{\gamma_2 n^{\kappa}}{\sqrt{\alpha_2 \log n + \beta_2}}(1 + Q_1(\log n) + o(1)). \tag{21}$$

(iii) [Double periodic case.] *If both matrices* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are logarithmically rationally related, then function* $Q_2(.)$ *is double periodic with small amplitude such that*

$$C_{n,n} = \frac{\gamma_2 n^{\kappa}}{\sqrt{\alpha_2 \log n + \beta_2}}(1 + Q_2(\log n) + o(1)). \tag{22}$$

In the three cases the constants γ_2, α_2 *and* β_2 *are explicitly computable.*

Remark In the double periodic case when the $\partial\mathcal{K}$ forms a lattice with commensurable vectors, the double periodic function reduces to a simple periodic function.

At last, we provide explicit expressions for some of constants in previously stated results, in particular in the most interesting Theorem 8. We denote $H(s_1, s_2) = 1 - \lambda(s_1, s_2)$. Let $H_1(s_1, s_2) = \frac{\partial}{\partial s_1}\lambda(s_1, s_2)$, $f(s_1, s_2) = \langle \boldsymbol{\pi}(s_1, s_2) | \mathbf{u}(s_1, s_2)\rangle$ and $g(s_1, s_2) = \langle \boldsymbol{\zeta}(s_1, s_2) | \mathbf{1}\rangle$.

Theorem 9. *Let* c_1 *and* c_2 *be between* -1 *and* 0. *In the general case we have*

$$C_{n,n} = n^{\kappa} \sum_{(s_1,s_2)\in\partial\mathcal{K}} \frac{\gamma(s_1,s_2) n^{-i\Im(s_1+s_2)}}{\sqrt{\alpha_2 \log n + \beta(s_1,s_2)}} \left(1 + O(\frac{1}{\log n})\right). \tag{23}$$

With H_1, α_2, $\beta_2(s_1, s_2)$ *and* $\gamma(s_1, s_2)$ *such*

$$\begin{aligned}
H_1 &= \frac{\partial}{\partial s_1} H(c_1, c_2) \\
\alpha_2 &= \frac{\Delta H - 2\frac{\partial^2}{\partial s_1 \partial s_2} H}{H_1}|_{(s_1,s_2)=(c_1,c_2)} \\
\beta_2(s_1, s_2) &= -\frac{\alpha_2}{2}\left(\Psi(s_1) + \Psi(s_2) + \frac{\frac{\partial}{\partial s_1} f + \frac{\partial}{\partial s_2} f}{f} + \frac{\frac{\partial}{\partial s_1} g + \frac{\partial}{\partial s_2} g}{g}\right) \\
&\quad + \Psi'(s_1) + \Psi'(s_2) + \frac{\Delta f - 2\frac{\partial^2}{\partial s_1 \partial s_2} f}{f} + \frac{\Delta g - 2\frac{\partial^2}{\partial s_1 \partial s_2} g}{g} \\
&\quad - \left(\frac{\frac{\partial}{\partial s_1} f - \frac{\partial}{\partial s_2} f}{f}\right)^2 - \left(\frac{\frac{\partial}{\partial s_1} g - \frac{\partial}{\partial s_2} g}{g}\right)^2 \\
&\quad + \frac{\alpha_2^2}{2} - \frac{\frac{\partial^3}{\partial s_1^3} H - \frac{\partial^3}{\partial s_1^2 \partial s_2} H - \frac{\partial^3}{\partial s_1 \partial s_2^2} H + \frac{\partial^3}{\partial s_2^3} H}{2H_1} + \left(\frac{\frac{\partial^2}{\partial s_1^2} H - \frac{\partial^2}{\partial s_2^2} H}{2H_1}\right)^2
\end{aligned}$$

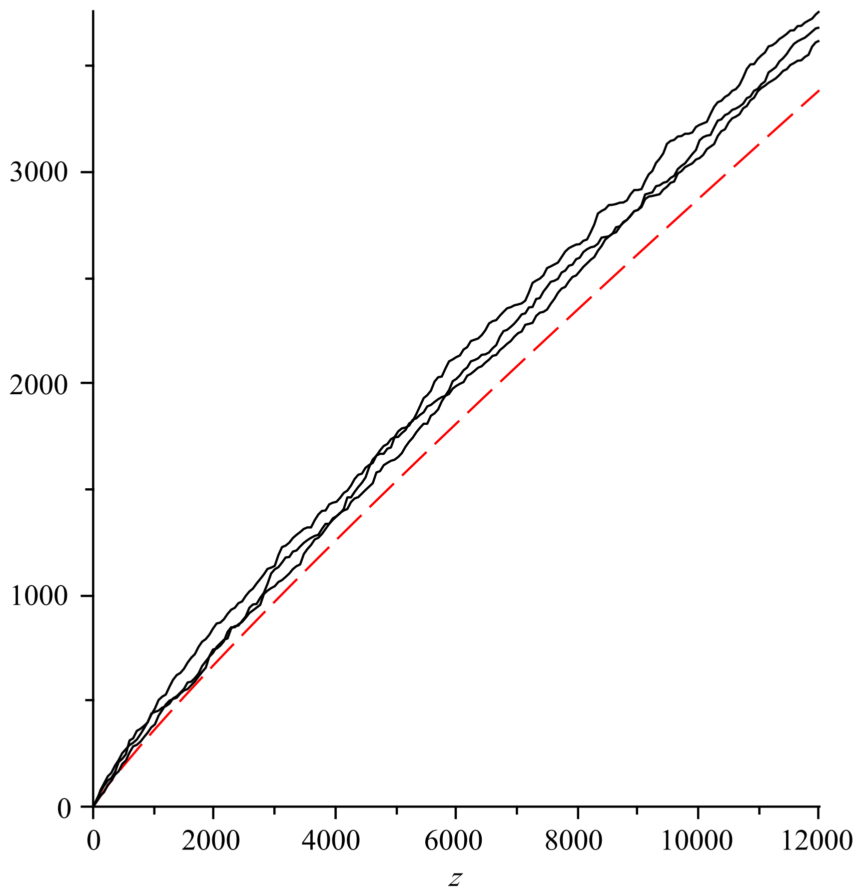

Figure 3: Joint complexity: three simulated trajectories (black) versus asymptotic average (dashed red) for the case $c_2 > 0$.

and

$$\gamma(s_1, s_2) = \frac{f(s_1)g(s_2)(s_1+1)\Gamma(s_1)(s_2+1)\Gamma(s_2)}{\lambda(s_2)\log|\mathcal{A}|\sqrt{2\pi}}.$$

The expression for α_2 seems to be asymmetric in (s_1, s_2). In fact, it is not since the maximum of $s_1 + s_2$ for $(s_1, s_2) \in \mathcal{K}$ attained on (c_1, c_2) necessarily implies that $\frac{\partial}{\partial s_1}H(c_1, c_2) = \frac{\partial}{\partial s_2}H(c_1, c_2)$. Formally the constant α_2 is equal to $\Delta\lambda_1(c_1, c_2)\langle \mathbf{1}|\nabla\lambda_1(c_1, c_2)\rangle$ where ∇ is the gradient operator, Δ the Laplacian operator $\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$.

Finally, we illustrate our results on two examples.

Example. In Figures 3 and 4 we plot the joint complexity for several pairs of strings X and Y. String X is generated by a Markov source with the transition matrix $\mathbf{P}$, and string Y is generated by a uniform memoryless source. We consider two Markov sources for X with the following transition matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{bmatrix}. \tag{24}$$

For the first $\mathbf{P}$ in Figure 3 we have $c_2 > 0$ (see Theorem 7) while for the second $\mathbf{P}$ in Figure 4 we have $c_2 < 0$ (cf. Theorem 4(i)).

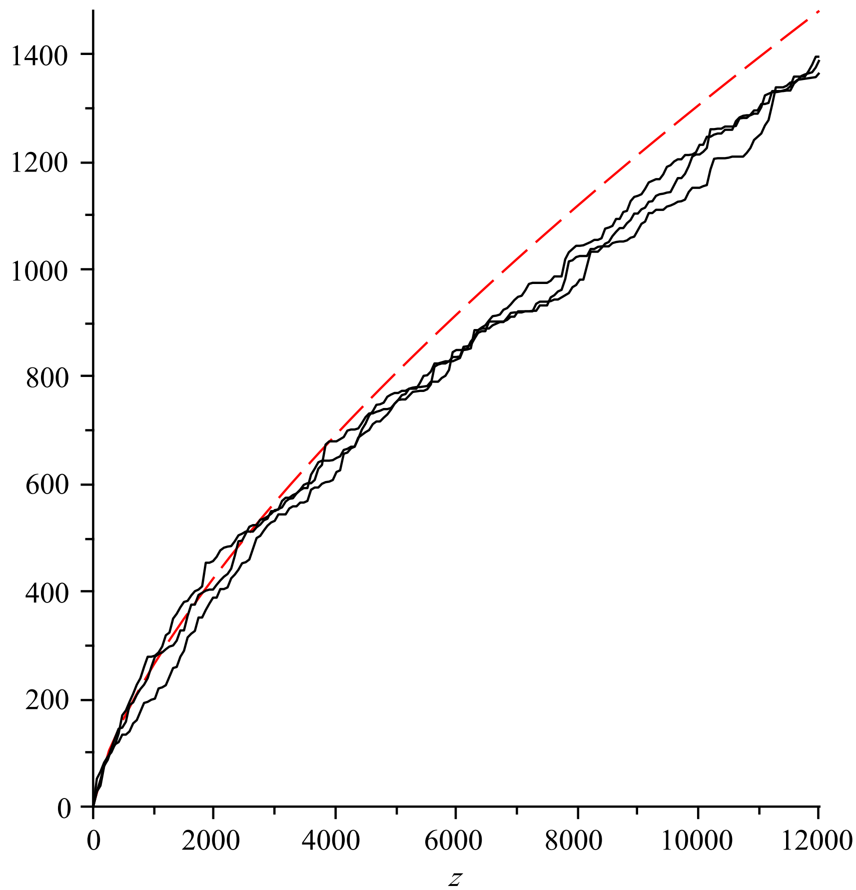

Figure 4: Joint semi-complexity: three simulated trajectories (black) versus asymptotic average (dashed red)for the case $c_2 < 0$.

3 Proof of Theorem 1

We recall that X and Y are two independent strings of length n and m, generated by two Markov sources characterized by transition matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively. In the previous section we write $|X|_w$ and $|Y|_w$ for the number of $w \in \mathcal{A}$ occurrences in X and Y. But it will be convenient to use another notation for these quantities, namely

$$O_n^1(w) := |X|_w, \quad O_m^2(w) := |Y|_w.$$

We shall use this notation interchangeably. Finally, we write $\mathcal{A}^+ = \mathcal{A}^* - \{\nu\}$, that is, for the set of all nonempty words. As observed in (1) we have

$$J_{n,m} = \sum_{w \in \mathcal{A}^+} P(O_n^1(w) \geq 1) P(O_m^2(w) \geq 1). \tag{25}$$

In [22, 10] the generating function of $P(O_n(w) \geq 1)$ for a Markov source is derived. It involves the *autocorrelation* polynomial of w, as discussed below. However, to make our analysis tractable we notice that $O_n^i(w) \geq 1$, $i = 1, 2$, is equivalent to w being a prefix of at least one of the n suffixes of X. But this is not sufficient to push forward our analysis. We need a second much deeper observation that replaces *dependent suffixes* with *independent strings* to shift analysis from suffix trees to tries, as already observed in [7] and briefly discussed in Section 2. In order to accomplish it, we consider two *sets* of infinite length strings of cardinality n and m, respectively, generated *independently* by Markov sources $\mathbf{P}_1$ and $\mathbf{P}_2$. As in Section 2 we denote by $\Omega_n^i(w)$ the number of strings for which w is a prefix when there are n strings generated by source i, for $i \in \{1, 2\}$. The average *joint prefix complexity* satisfies (2) that we repeat below

$$C_{n,m} = \sum_{w \in \mathcal{A}^+} P(\Omega_n^1(w) \geq 1) P(\Omega_m^2(w) \geq 1). \tag{26}$$

Before we prove our first main result Theorem 1 we need some preliminary work. First, observe that it is relatively easy to compute the probability $P(\Omega_n^i(w) \geq 1)$. Indeed,

$$P(\Omega_n^i(w) \geq 1) = 1 - (1 - P_i(w))^n.$$

Notice that the quantity $1 - (1 - P(w))^n$ is the probability that $|\mathcal{X}_n|_w > 0$ where $\mathcal{X}_n$ is a set of n independently generated strings. To prove Theorem 1 we must show that for Markov sources when $n \to \infty$

$$P(O_n^i(w) \geq 1) \sim P(\Omega_n^i(w) \geq 1)$$

which we do in the next key lemma.

We denote by $\mathcal{B}_k$ the set of words of length k such that a word $w \in \mathcal{B}_k$ does not overlap with itself over more than $k/2$ characters (see [10, 7, 20] for more precise definition). It is proved in [20] that

$$\sum_{w \in \mathcal{A}^k - \mathcal{B}_k} P(w) = O(\delta_1^k)$$

where $\mathcal{A}^k$ is the set of all words of length k and $\delta_1 < 1$ is the largest element of the Markovian transition matrix $\mathbf{P}$. In order to allow some transition probabilities to be equal to 0 we define

$$\begin{aligned} p &= \exp\left(\lim \sup_{k, w \in \mathcal{A}^k} \frac{\log P(w)}{k}\right) \\ q &= \exp\left(\lim \inf_{k, w \in \mathcal{A}^k, P(w) \neq 0} \frac{\log P(w)}{k}\right). \end{aligned}$$

These quantities exist and are smaller than 1 since $\mathcal{A}$ is a finite alphabet [21, 23]. In fact, they are related to Renyi's entropy of order $\pm\infty$, respectively [23]. We write $\delta = \sqrt{p} < 1$.

Let X_n be a string of length n generated by a Markov source. For $w \in \mathcal{A}^*$, define

$$d_n(w) = P(O_n(w) > 0) - (1 - (1 - P(w))^n). \tag{27}$$

We prove the following lemma

Lemma 3. *Let $w \in \mathcal{A}^k$ be of length k. There exists $\rho > 1$ and a sequence $R_n(w) = O(P(w)\rho^{-n})$ such that for all $1 > \epsilon > 0$ we have:*
(i) *for $w \in \mathcal{B}_k$: $d_n(w) = O((nP(w))^{\epsilon} k\delta^k) + R_n(w)$;*
(ii) *for $w \in \mathcal{A}^k - \mathcal{B}^k$: $d_n(w) = O((nP(w))^{\epsilon}\delta^k) + R_n(w)$.*

Proof. Let $N_0(z) = \sum_{n\geq 0} P(O_n(w) = 0)z^n$. We know from [10] that

$$N_0(z) = \frac{S_w(z)}{D_w(z)}$$

where $S_w(z)$ is the autocorrelation polynomial of word w and $D_w(z)$ is defined as follows

$$D_w(z) = S_w(z)(1-z) + z^k P(w)\left(1 + F_{w_1,w_k}(z)(1-z)\right), \tag{28}$$

where $k = |w|$ is the length of word w with first symbol w_1 and the last symbol w_k. Here $F_{a,b}(z)$ for $(a,b) \in \mathcal{A}^2$ is a generating function that depends on the Markov sources, as describe below. We also write $F_w(z)$ when $w_1 = a$ and $w_k = b$.

Let $\mathbf{P}$ be the transition matrix of the Markov source. Let $\boldsymbol{\pi}$ be its stationary vector and for $a \in \mathcal{A}$ let π_a be its coefficient at symbol a. The vector $\mathbf{1}$ is the vector with all coefficients equal to 1 and $\mathbf{I}$ is the identity matrix. Assuming that the symbol a (resp. b) is the first (resp. last) character of w, we have [22]

$$F_w(z) := F_{a,b(z)} = \frac{1}{\pi_a}\left[(\mathbf{P} - \boldsymbol{\pi}\otimes\mathbf{1})\left(\mathbf{I} - z(\mathbf{P} + \boldsymbol{\pi}\otimes\mathbf{1})\right)^{-1}\right]_{a,b} \tag{29}$$

where $\otimes$ is the tensor product, and $[A]_{a,b}$ denotes the (a,b)th coefficient of matrix A. An alternative way to express $F_w(z)$ is

$$F_w(z) = \frac{1}{\pi_a}\langle \mathbf{e}_a(\mathbf{P} - \boldsymbol{\pi}\otimes\mathbf{1})\left(\mathbf{I} - z(\mathbf{P} + \boldsymbol{\pi}\otimes\mathbf{1})\right)^{-1}\mathbf{e}_b\rangle \tag{30}$$

where $\mathbf{e}_c$ for $c \in \mathcal{A}$ is the vector with a 1 at the position corresponding to symbol c and all other coefficients are 0.

By the spectral representation of matrix $\mathbf{P}$ we have [14]

$$\mathbf{P} = \boldsymbol{\pi}\otimes\mathbf{1} + \sum_{i>1}\lambda_i \mathbf{u}_i \otimes \boldsymbol{\zeta}_i$$

where λ_i for $i \geq 1$ is the ith eigenvalue (in decreasing order) of matrix $\mathbf{P}$ (with $\lambda_1 = 1$), and $\mathbf{u}_i$ (resp. $\boldsymbol{\zeta}_i$) are their corresponding right (resp. left) eigenvectors. Thus

$$(\mathbf{P} - \boldsymbol{\pi}\otimes\mathbf{1})\left(\mathbf{I} - z(\mathbf{P} + \boldsymbol{\pi}\otimes\mathbf{1})\right)^{-1} = \sum_{i>1}\frac{\lambda_i}{1-\lambda_i z}\mathbf{u}_i \otimes \boldsymbol{\zeta}_i \tag{31}$$

and therefore the function $F_w(z)$ is defined for all z such that $|z| < \frac{1}{|\lambda_2|}$ and is uniformly $O(\frac{1}{1-|\lambda_2 z|})$.

We now follow the approach from [20] that extends to Markovian sources the analysis presented in [7] for memoryless sources (see also [10]). Let

$$\Delta_w(z) = \sum_{n\geq 0} d_n(w) z^n$$

be the generating function of $d_n(w)$ defined in (27). After some algebra we arrive at

$$\Delta_w(z) \quad = \quad \frac{P(w)z}{1-z}\left(\frac{1+(1-z)F_w(z)}{D_w(z)} - \frac{1}{1-z+P(w)z}\right). \tag{32}$$

We have

$$d_n(w) = \frac{1}{2i\pi}\oint \Delta_w(z)\frac{dz}{z^{n+1}},$$

integrated on any loop encircling the origin in the definition domain of $d_w(z)$. Extending the result from [7], the authors of [20] show that there exists $\rho > 1$ such that the function $D_w(z)$ defined in (28) has a single root in the disk of radius ρ. Let A_w be this root. We have via the residue formula

$$d_n(w) = \mathrm{Res}(\Delta_w(z), A_w)A_w^{-n} - (1-P(w))^n + d_n(w,\rho) \tag{33}$$

where $\mathrm{Res}(f(z), A)$ denotes the residue of function $f(z)$ on complex number A. Thus

$$d_n(w,\rho) = \frac{1}{2i\pi}\oint_{|z|=\rho} \Delta_w(z)\frac{dz}{z^{n+1}}. \tag{34}$$

We have

$$\mathrm{Res}(\Delta_w(z), A_w) = \frac{P(w)\,(1+(1-A_w)F_w(A_w))}{(1-A_w)C_w} \tag{35}$$

where $C_w = D'_w(A_w)$. But since $D_w(A_w) = 0$ we can write

$$\mathrm{Res}(\Delta_w(z), A_w) = -\frac{A_w^{-k}S_w(A_w)}{C_w}. \tag{36}$$

We can take the asymptotic expansion of A_w and C_w as it is described in [10], in Lemma 8.1.8 and Theorem 8.2.2. Anyhow the expansions were done in the memoryless case. But an extension to Markov sources simply consists in replacing $S_w(1)$ into $S_w(1) + P(w)F_w(1)$, so we find

$$\left\{\begin{array}{lcl} A_w & = & 1 + \frac{P(w)}{S_w(1)} \\ & & +P^2(w)\left(\frac{k-F_w(1)}{S_w^2(1)} - \frac{S'_w(1)}{S_w^3(1)}\right) + O(P(w)^3), \\ C_w & = & -S_w(1) + P(w)\left(k - F_w(1) - 2\frac{S'_w(1)}{S_w(1)}\right) \\ & & +O(P(w)^2). \end{array}\right. \tag{37}$$

These expansions also appear in [20].

From now the proof takes the same path as the proof of Theorem 8.2.2 in [10]. We define the function

$$d_w(x) = \frac{A_w^{-k}S_w(A_w)}{C_w}A_w^{-x} - (1-P(w))^x. \tag{38}$$

More precisely, we define the function $\bar{d}_w(x) = d_w(x) - d_w(0)e^{-x}$. Its Mellin transform [23] is

$$d_w^*(s)\Gamma(s) = \int_0^\infty \bar{d}_w(x)x^{s-1}dx$$

defined for all $\Re(s) \in (-1, 0)$ with

$$d_w^*(s) \;=\; \frac{A_w^{-k}S_w(A_w}{C_w}((\log A_w)^{-s} - 1) + 1 - (-\log(1-P(w))))^{-s} \tag{39}$$

where $\Gamma(s)$ is the Euler gamma function. When $w \in \mathcal{B}_k$ with the expansion of A_w and since $S_w(1) = 1 + O(\delta^k)$ and $S'_w(1) = O(k\delta^k)$, we find that similarly as in [10]

$$d_w^*(s) = O(|s|k\delta^k)P(w)^{1-s} \tag{40}$$

and therefore by the reverse Mellin transform, for all $1 > \epsilon > 0$:

$$\begin{aligned} \bar{d}_w(n) &= \frac{1}{2i\pi}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} d_w^*(s)\Gamma(s)s^{-n}ds \\ &= O(n^\epsilon P(w)^\epsilon k\delta^k). \end{aligned} \tag{41}$$

When $w \in \mathcal{A}^k - \mathcal{B}_k$ it is not true that $S_w(1) = 1 + O(\delta^k)$, thus it is shown in [20] that there exists $\alpha > 0$ such that for all $w \in \mathcal{A}^*$: $S_w(z) > \alpha$ for all z such that $|z| \le \rho$. Therefore we get $\bar{d}_w(n) = O(n^\epsilon P(w)^\epsilon)$.

We set

$$R_n(w) = d_w(0)e^{-n} + d_n(w,\rho). \tag{42}$$

We first investigate the quantity $d_w(0)$. We prove that $d_w(0) = O(P(w))$. Noticing that

$$S_w(A_w) = S_w(1) + \frac{P(w)}{S_w(1)} S'_w(1) + O(P(w)^2)$$

we have the expansion

$$-\frac{A_w^{-k} S_w(A_w)}{C_w} = 1 - \frac{P(w)}{S_w(1)} \left(F_w(1) + \frac{S'_w(1)}{S_w(1)} \right) + O(P(w)^2). \tag{43}$$

Thus

$$d_w(0) = -\frac{P(w)}{S_w(1)} \left(F_w(1) + \frac{S'_w(1)}{S_w(1)} \right) + O(P(w))^2). \tag{44}$$

Thus $d_w(0) = O(P(w))$.

Now we need to consider $d_n(w,\rho)$. Since $\Delta_n(z)$ is clearly $O(P(w))$ and the integral $\oint \Delta(z) \frac{dz}{z^{n+1}}$ is over the circle of radius, the result is $O(P(w)\rho^{-n})$. □

Now we are ready to prove Theorem 1.

Proof. Again staring with

$$J_{n,m} = \sum_{w \in \mathcal{A}^*} P(O_n^1(w) > 0) P(O_m^2(w) > 0) \tag{45}$$

we note that

$$\begin{aligned} P(O_n^1(w) > 0)) &= 1 - (1 - P_1(w))^n + d_n^1(w), \\ P(O_m^2(w) > 0) &= 1 - (1 - P_2(w))^m + d_m^2(w). \end{aligned}$$

Thus

$$\begin{aligned} J_{n,m} &= C_{n,m} + \sum_{w \in \mathcal{A}} d_n^1(w) P(O_m^2(w) > 0) \\ &+ \sum_{w \in \mathcal{A}} (1 - (1 - P_1(w))^n) d_m^2(w). \end{aligned} \tag{46}$$

We will give the proof for the first sum, since the proof for the second proof being somewhat similar. When $w \in \mathcal{B}_k$, we have for all $\epsilon > 0$

$$d_n^1(w) = O(n^\epsilon P_1(w)^\epsilon k \delta_1^k) + R_n^1(w) \tag{47}$$

and the $R_n^1(w)$ terms are all $O(P_1(w)\rho^{-n})$. We look at the sum

$$\sum_k \sum_{w \in \mathcal{B}_k} n^\epsilon P(w)^\epsilon k \delta_1^k.$$

It is smaller than

$$\sum_k \sum_{w \in \mathcal{A}^k} n^\epsilon P_1(w) k (q^{\epsilon - 1} \delta_1)^k$$

which is equal to $n^\epsilon \sum_k k(q^{\epsilon-1}\delta_1)^k$. By choosing a value ϵ_1 of ϵ enough close to 1 so that $q^{\epsilon_1 - 1}\delta_1 < 1$ we have the n^{ϵ_1} order. Notice that with $\delta_1 = \sqrt{p}$ we must conclude that $1/2 < \epsilon_1 < 1$.

When $w \in \mathcal{A}^k - \mathcal{B}_k$ the δ_1^k factor disappears in the right-hand side expression for $d_n^1(w)$. But in this case

$$\sum_{w \in \mathcal{A}^k - \mathcal{B}_k} n^\epsilon P(w) (q^{\epsilon-1})^k = O(n^\epsilon \delta_1^k (q^{\epsilon-1})^k),$$

and we conclude similarly.

It remains the sum $\sum_{w\in\mathcal{A}^*} R_n^1(w)P_2(O_m(w)>0)$. For this we remark that $P_2(O_m(w)>0)=O(mP_2(w))$. Therefore the sum is of order $\rho^{-n}m\sum_{w\in\mathcal{A}^*}P_1(w)P_2(w)$. It turns out that

$$\sum_{w\in\mathcal{A}^k} P_1(w)P_2(w)=O(\lambda_{12}^k)$$

where λ_{12} is the main eigenvalue of the Shur product matrix $\mathbf{P}_1\star\mathbf{P}_2$ (also known denoted $\mathbf{P}(-1,-1)$. Since $\lambda_{12}<1$ the sum converges and is $O(\rho^{-n}m)$. $\square$

4 Some Preliminary Results

In this section we first derive the recurrence on $C_{n,m}$ which will lead to the functional equation on the double Poisson transform $C(z_1,z_2)$ of $C_{n,m}$, that in turn allows us to find the double Mellin transform $C^*(s_1,s_2)$ of $C(z_1,z_2)$. Finally applying a double depoissonization we first recover the original $C_{n,n}$ and ultimately the joint string complexity $J_{n,n}$ through Theorem 1.

4.1 Functional Equations

Let $a\in\mathcal{A}$ and define

$$C_{a,m,n}=\sum_{w\in a\mathcal{A}*} P(\Omega_n^1(w)\geq 1)P(\Omega_m^2(w)\geq 1)$$

where $w\in a\mathcal{A}^*$ means that w starts with an $a\in\mathcal{A}$. We recall that $\Omega_n^i(w)$ represents the number of strings of length n that start with prefix w.

Notice that $C_{a,m,n}=0$ when $n=0$ or $m=0$. Using Markov nature of the string generation, the quantity $C_{a,n,m}$ for $n,m\geq 1$ satisfies the following recurrence for all $a,b\in\mathcal{A}$

$$\begin{aligned}C_{b,n,m} &= 1+\sum_{a\in\mathcal{A}}\sum_{n_a,m_a}\binom{n}{n_a}\binom{m}{m_a}\\ &\quad\times(P_1(a|b))^{n_a}(1-P_1(a|b))^{n-n_a}\\ &\quad\times(P_2(a|b))^{m_a}(1-P_2(a|b))^{m-m_a}C_{a,n_a,m_a},\end{aligned}$$

where n_a (resp. m_a) denotes the number of strings among n (resp. m) independent strings from source 1 (resp. 2) that have symbol a followed by symbol b. Indeed, partitioning $b\mathcal{A}^*$ as $\{b\}+\sum_{a\in\mathcal{A}}a\mathcal{A}^*$ we obtain the recurrence noting that strings are independent and $\binom{n}{n_a}(P_1(a|b))^{n_a}(1-P_1(a|b))^{n-n_a}$ is the probability of n_a out of n starting with ba. starts

In a similar fashion, the *unconditional* average $C_{n,m}$ satisfies for $n,m\geq 2$

$$C_{n,m}=1+\sum_{a\in\mathcal{A}}\sum_{n_a,m_a}\binom{n}{n_a}\binom{m}{m_a}\pi_1^{n_a}(a)(1-\pi_1(a))^{n-n_a}$$

$$\times\pi_2^{m_a}(a)(1-\pi_2(a))^{m-m_a}C_{a,n,m}.$$

To solve it we introduce the double Poisson transform of $C_{a,n,m}$ as

$$C_a(z_1,z_2)=\sum_{n,m\geq 0}C_{a,n,m}\frac{z_1^nz_2^m}{n!m!}e^{-z_1-z_2} \tag{48}$$

that translates the above recurrence into the following functional equation:

$$\begin{aligned}C_b(z_1,z_2) &= (1-e^{-z_1})(1-e^{-z_2})\\ &\quad+\sum_{a\in\mathcal{A}}C_a\left(P_1(a|b)z_1,P_2(a|b)z_2\right).\end{aligned} \tag{49}$$

To simplify it, we define the double Poisson transform

$$C(z_1, z_2) = \sum_{n,m\geq 0} C_{n,m} \frac{z_1^n z_2^m}{n!m!} e^{-z_1-z_2} \tag{50}$$

finding that

$$\begin{aligned} C(z_1, z_2) &= (1-e^{-z_1})(1-e^{-z_2}) \\ &+ \sum_{a\in\mathcal{A}} C_a(\pi_1(a)z_1, \pi_2(a)z_2). \end{aligned} \tag{51}$$

Our goal now is to find asymptotic expansion of $C(z_1, z_2)$ as $z_1, z_2 \to \infty$ in a cone around the real axis. This will be accomplished in the next subsection using double Mellin transform. Granted it, we shall appeal to a double depoissonization result to recover asymptotically $C_{n,m}$ and ultimately $J_{n,m}$.

4.2 Double DePoissonization

Once we know $C(z_1, z_2)$ for $z_1 = n, z_2 = m \to \infty$ we then need to recover $C_{n,m}$. Double depoissonization lemma discussed and proved in [10] (see Lemma 10.3.4) allows us to do exactly that but in order to apply it we need to postulate some conditions on the underlying Poisson transforms. We briefly review double depoissonization next.

For a double sequence $a_{n,m}$ define

$$\begin{aligned} f(z_1, z_2) &= \sum_{m,n=0}^{\infty} a_{n,m} \frac{z_1^n}{n!} e^{-z_1} \frac{z_2^m}{m!} e^{-z_2}, \\ f_n(z_2) &= \sum_{m=0}^{\infty} a_{n,m} \frac{z_2^m}{m!} e^{-z_2}. \end{aligned}$$

We notice that $f(z_1, z_2)$ is the Poisson transform of the sequence $f_n(z_2)$ with respect to the variable z_1. Now we postulate certain conditions on $f(z_1, z_2)$ and $f_n(z_2)$ that will allow us to extract asymptotics of $a_{n,m}$ from $f(z_1, z_2)$.

First depoissonization. For $z_2 \in \mathcal{S}_\theta := \{z_2 : \ \arg(z_2) < \theta\}$ we postulate that there exist constants β, α, B and D such that

$$\begin{aligned} z_1 \in \mathcal{S}_\theta : |f(z_1, z_2)| &< B(|z_1|^\beta + |z_2|^\beta) \\ z_1 \notin \mathcal{S}_\theta : |f(z_1, z_2)e^{z_1}| &< D|z_2|^\beta e^{\alpha|z_1|}. \end{aligned}$$

Therefore, from the one-dimensional analytic depoissonization of [8, 23] for $z_2 \in \mathcal{S}_\theta$, we have for all integers $k > 0$

$$f_n(z_2) = f(n, z_2) + O\left(n^{\beta-1} + \frac{|z_2|^\beta}{n}\right) + O(|z_2|^\beta n^{\beta-k}).$$

Similarly, when $z_2 \notin \mathcal{S}_\theta$ we postulate

$$\begin{aligned} z_1 \in \mathcal{S}_\theta : |f(z_1, z_2)e^{z_2}| &< D|z_1|^\beta e^{\alpha|z_2|} \\ z_1 \notin \mathcal{S}_\theta : |f(z_1, z_2)e^{z_1+z_2}| &< De^{\alpha|z_1|+\alpha|z_2|}. \end{aligned}$$

Thus for all integer k and $\forall z_2 \notin \mathcal{S}_th$

$$f_n(z_2)e^{z_2} = f(n, z_2)e^{z_2} + O(n^{\beta-1}e^{\alpha|z_2|}) + O(n^{\beta-k}e^{\alpha|z_2|}).$$

Second depoissonization. The above two conditions on $f_n(z_2)$, respectively for $z_2 \in \mathcal{S}_\theta$ and $z_2 \notin \mathcal{S}_\theta$, allow us to depoissonize $f_n(z_2)$. For all $k > \beta$:

- for $z_2 \in \mathcal{S}_\theta$: $f_n(z_2) = O(n^\beta + |z_2|^\beta)$;

- for $z_2 \notin \mathcal{S}_\theta$: $f_n(z_2)e^{z_2} = O(n^\beta e^{\alpha|z_2|})$.

These estimates are uniform. Therefore,

$$a_{n,m} = f_n(m) + O\left(\frac{n^\beta}{m} + \frac{m^\beta}{n}\right) + O\left(n^\beta m^{\beta-k}\right).$$

Since

$$f_n(m) = f(n,m) + O\left(n^{\beta-1} + \frac{m^\beta}{n}\right)$$

and setting $k > \beta + 1$, we find the desired estimate.

Now we are ready for formulate our depoissonization lemma. In [10] it is shown that $C(z_1, z_2)$ satisfies depoissonization conditions for memoryless sources. In Appendix we prove the following lemma.

Lemma 4 (DePoissonization)**.** *We have*

$$C_{n,m} = C(n,m) + O\left(\frac{n}{m} + \frac{m}{n}\right)$$

for large n and m.

To find $C(n,m)$ from $C(z_1, z_2)$ we follow the Mellin transform approach, however for general sources we need to consider a double Mellin transform. We start in the next subsection with a simple case when $\mathbf{P}_1 = \mathbf{P}_2$.

4.3 Mellin Transform for $\mathbf{P}_1 = \mathbf{P}_2$: Proof of Theorem 2

We first present a general result for identical Markov sources, that is, $\mathbf{P}_1 = \mathbf{P}_2 = \mathbf{P}$ proving Theorem 2. In this case (49) can be rewritten with $c_a(z) = C_a(z,z)$:

$$c_b(z) = (1 - e^{-z})^2 + \sum_{a \in \mathcal{A}} c_a\left(P(a|b)z\right). \tag{52}$$

This equation is directly solvable by the Mellin transform defined as

$$c_a^*(s) = \int_0^\infty c_a(x) x^{s-1} dx$$

that exists in the fundamental strip $-2 < \Re(s) < -1$. Properties of Mellin transform can be found in [4, 23]. It follows that for all $b \in \mathcal{A}$ [23]

$$c_b^*(s) = (2^{-s} - 2)\Gamma(s) + \sum_{a \in \mathcal{A}} (P(a|b))^{-s} c_a^*(s). \tag{53}$$

It is better to write it in the matrix form. Let $\mathbf{c}(s) = [c_a^*(s)]_{a \in \mathcal{A}}$ be the vector of Mellin transforms $c_a^*(s)$ and and $\mathbf{P}(s) = [P^{-s}(a|b)]_{a,b \in A}$. Then (53) becomes

$$\mathbf{c}(s) = (2^{-s} - 2)\mathbf{1}(\mathbf{I} - \mathbf{P}(s))^{-1}$$

where, again, $\mathbf{1}$ is the vector of dimension $|\mathcal{A}|$ made of all 1's, and $\mathbf{I}$ is the identity matrix.

We now can derive the Mellin transform $c^*(s)$ of $c(z) := C(z,z)$ representing the unconditional joint string complexity. From above and (51) we arrive at

$$c^*(s) = (2^{-s} - 2)\Gamma(s) + \sum_{a \in \mathcal{A}} (\pi(a))^{-s} c_a^*(s)\ .$$

which in matrix form can be rewritten as

$$c^*(s) = (2^{-s} - 2)\Gamma(s)\left(1 + \langle \mathbf{1}(\mathbf{I} - \mathbf{P}(s))^{-1} | \boldsymbol{\pi}(s) \rangle\right) \tag{54}$$

where $\boldsymbol{\pi}(s)$ is the the vector made of coefficients $\pi(a)^{-s}$ and we recall $\langle \mathbf{x}|\mathbf{y}\rangle$ is the inner product of vectors $\mathbf{x}$ and $\mathbf{y}$.

To find the behavior of $c(z)$ for large z near the real axis we apply the inverse Mellin approach as discussed in [4, 23]. We observe that

$$c(z) = \frac{1}{2\pi i}\int_{-3/2-\infty}^{-3/2+\infty} c^*(s)z^{-s}ds.$$

The asymptotics of $c(z)$ for $|\arg(z)| < \theta$ is given by the residues of the function $c^*(s)z^{-s}$ occurring at the poles $s = -1$ and $s = 0$. They are respectively equal to

$$\frac{2\log 2}{h}z$$

and

$$-1 - \langle \mathbf{1}(\mathbf{I} - \mathbf{P}(0,0))^{-1}\boldsymbol{\pi}(0)\rangle.$$

The first residues comes from the singularity of $(\mathbf{I} - \mathbf{P}(s))^{-1}$ at $s = -1$. This leads to Theorem 2(i). When $\mathbf{P}$ is rationally related then there are additional poles on a countable set of complex numbers s_k regularly spaced on the line $\Re(s_k) = -1$, and such that $\mathbf{P}(s_k)$ has eigenvalue 1. These poles contributes to the periodic terms of Theorem 2(ii). The proof of Theorem 2 is now complete.

4.4 Double Mellin transform: Case $\mathbf{P}_1 \neq \mathbf{P}_2$

From now on we only consider the case $\mathbf{P}_1 \neq \mathbf{P}_2$, and therefore we need to study properties of $C_a(z,z)$ through double Mellin transform defined as

$$C_a^*(s_1,s_2) = \int_0^\infty\int_0^\infty C_a(z_1,z_2)z_1^{s_1-1}z_2^{s_2-1}dz_1dz_2 \tag{55}$$

or similarly the Mellin transform $\mathbf{C}^*(s_1,s_2)$ applied to $\mathbf{C}(z_1,z_2)$, provided we can find a strip $a < \Re(s_1), \Re(s_2) < b$ where the above transforms exist. Since for any $y \in \mathbb{R}^+$ and a function f we have the identity

$$\int_0^\infty f(yx)x^{s-1}dx = a^{-s}\int_0^\infty f(x)x^{s-1}dx$$

we conclude that

$$\mathbf{C}^*(s_1,s_2) = \Gamma(s_1)\Gamma(s_2)\mathbf{1} + \mathbf{P}(s_1,s_2)\mathbf{C}^*(s_1,s_2) \tag{56}$$

or formally

$$\mathbf{C}^*(s_1,s_2) = \Gamma(s_1)\Gamma(s_2)\left(\mathbf{I} - \mathbf{P}(s_1,s_2)\right)^{-1}\mathbf{1}. \tag{57}$$

However, the above formal derivation needs to be amended with a careful analysis of the convergence issues, which we do next. Notice that for any $a \in \mathcal{A}$: $C_a(z_1,z_2) = O(|z_1| + |z_2|)$ when $z_1, z_2 \to \infty$. But as easy to check $C_a(z_1,z_2)$ is also $O(|z_1| + |z_2|)$ when $z_1, z_2 \to 0$, therefore the Mellin transform is not appropriately defined in (55). To correct it, we now introduce correction terms in the expression of $C_a(z_1,z_2)$ so that the corresponding Mellin transform exists for $-2 < \Re(s_1), \Re(s_2) < -1$.

To continue, we now define a slightly modified Mellin transform, namely

$$\widetilde{\mathbf{C}}(z_1,z_2) = \mathbf{C}(z_1,z_2) - \mathbf{D}(z_1,z_2)$$

where

$$\mathbf{D}(z_1,z_2) = z_1e^{-z_1}\mathbf{C}_1(z_2) + z_2e^{-z_2}\mathbf{C}_2(z_1) - \mathbf{C}_{1,1}z_1z_2e^{-z_1-z_2}$$

with

$$\begin{aligned}
\mathbf{C}_1(z) &= \frac{\partial}{\partial z_1}\mathbf{C}(z_1,z_2)|_{(z_1,z_2)=(0,z)}\\
\mathbf{C}_2(z) &= \frac{\partial}{\partial z_2}\mathbf{C}(z_1,z_2)|_{(z_1,z_2)=(z,0)}.
\end{aligned}$$

Notice that $\widetilde{\mathbf{C}}(z_1, z_2)$ is now $O(|z_1|^2+|z_2|^2)$ when $z_1, z_2 \to 0$. We can show that $\mathbf{C}(z_1, z_2) = O(|z_1|+|z_2|)$ for (z_1, z_2) in four dimension cone containing $\mathbb{R}^+ \times \mathbb{R}^+$, therefore by Ascoli theorem $\frac{\partial}{\partial z_1}\mathbf{C}(z_1, z_2) = O(1)$ in the same cone, $D_1(z)$ is $O(|z|)$ for $z \in \mathbb{R}^+$, and similarly $D_2(z)$ is $O(1)$. All of this is to state that $\widetilde{\mathbf{C}}(z_1, z_2)$ is $O(|z_1| + |z_2|)$ when $z_1, z_2 \to \infty$, thus the Mellin transform of $\widetilde{\mathbf{C}}(z_1, z_2)$ is well defined for $\Re(s_1), \Re(s_2) \in (-2, -1)$. Let $\widetilde{\mathbf{C}}^*(s_1, s_2)$ be the corresponding Mellin transform.

For $a \in \mathcal{A}$ let $C_{1a}(z)$ be the coefficient of the vector $\mathbf{C}_1(z)$ corresponding to the symbol a. For $b \in \mathcal{A}$ we have the functional equation

$$C_{1b}(z) = 1 - e^{-z} + \sum_{a\in\mathcal{A}} P_1(a|b)C_{1a}(P_2(a|b)z) \tag{58}$$

and the Mellin transform of $\mathbf{C}_1(z)$, say $\mathbf{C}_1^*(s)$ formally satisfies

$$\mathbf{C}_1(s) = -\Gamma(s)\mathbf{1} + \mathbf{P}(-1, s)\mathbf{C}_1(s) \tag{59}$$

or

$$\mathbf{C}_1(s) = -\Gamma(s)(\mathbf{I} - \mathbf{P}(-1, s))^{-1}\mathbf{1}. \tag{60}$$

Similarly the Mellin transform $\mathbf{C}_2(s)$ of $\mathbf{C}_2(z)$ satisfies $\mathbf{C}_2(s) = \Gamma(s)(\mathbf{I} - \mathbf{P}(s, -1))^{-1}\mathbf{1}$. To finish, we notice that $\mathbf{C}_{1,1} = (\mathbf{I} - \mathbf{P}(-1, -1))^{-1}$.

Denoting $\widehat{\mathbf{C}}(s_1, s_2) = (\mathbf{I} - \mathbf{P}(s_1, s_2))^{-1}\mathbf{1}$, we find

$$\begin{aligned}\widetilde{\mathbf{C}}^*(s_1, s_2) \quad &= \quad \Gamma(s_1)\Gamma(s_2)\Big(\widehat{\mathbf{C}}(s_1, s_2) + s_2\widehat{\mathbf{C}}(s_1, -1) \\ &\quad + s_1\widehat{\mathbf{C}}(-1, s_2) + s_1 s_2\widehat{\mathbf{C}}(-1, -1)\Big)\end{aligned}$$

finally leading to

$$C^*(s_1, s_2) = \Gamma(s_1)\Gamma(s_2)\left(1 + \langle \boldsymbol{\pi}(s_1, s_2)|\widehat{\mathbf{C}}(s_1, s_2)\rangle\right) \tag{61}$$

where $\boldsymbol{\pi}(s_1, s_2)$ denotes the vector composed of $\pi_1(a)^{-s_1}\pi_2(a)^{-s_2}$ for $a \in \mathcal{A}$ and $\langle|\rangle$ is the vector internal product.

Our goal is to find $C_{n,n}$ (i.e., $n = m$). But by depoissonization it is asymptotically equal to $C(n, n)$, therefore we must find $\mathbf{C}(z, z)$ which by the inverse Mellin transform becomes

$$\mathbf{C}(z, z) \quad = \quad \mathbf{D}(z, z) + \frac{1}{(2i\pi)^2}\int_{\rho_1}\int_{\rho_2}\widetilde{\mathbf{C}}^*(s_1, s_2)z^{-s_1-s_2}ds_1ds_2.$$

After some algebra we finally arrive at

$$C(z, z) \quad = \quad (1 - e^{-z})^2 + \tag{62}$$

$$\frac{1}{(2i\pi)^2}\int_{\rho_1}\int_{\rho_2}\Gamma(s_1)\Gamma(s_2)\langle\boldsymbol{\pi}(s_1, s_2)|\widehat{\mathbf{C}}(s_1, s_2)+s_2\widehat{\mathbf{C}}(s_1, -1)+s_1\widehat{\mathbf{C}}(-1, s_2)+s_1s_2\widehat{\mathbf{C}}(-1, -1)\rangle z^{-s_1-s_2}ds_1ds_2$$

where the integration is over the lines $\Re(s_1) = \rho_1$ and $\Re(s_1) = \rho_2$ with (ρ_1, ρ_2) belonging to the fundamental strip of $C^*(s_1, s_2)$: $(-2, -1)$. We shall analyze asymptotically (62) in the next sections.

4.5 Properties of the Kernel

We recall from Section 2 that we define the kernel $\overline{\mathcal{K}}$ as the set of complex tuples (s_1, s_2) such that $\mathbf{P}(s_1, s_2)$ has largest eigenvalue $\lambda(s_1, s_2) = 1$. Furthermore, we also define $\partial\mathcal{K}$ as the subset of $\overline{\mathcal{K}}$ consisting of the pairs (s_1, s_2) such $\Re(s_1, s_2) = (c_1, c_2)$ where

$$(c_1, c_2) = \arg\min_{(s_1,s_2)\in\mathcal{K}}\{-s_1 - s_2\}.$$

We also denote $\partial\mathcal{K}^* = \partial\mathcal{K} - \{(c_1, c_2)\}$.

Let us start with the structure of the set $\partial\mathcal{K}$.

Definition 6. *Let* $\mathbf{P}$ *be a matrix on* $\mathcal{A} \times \mathcal{A}$ *of complex coefficients* p_{ab} *for all* $(a, b) \in \mathcal{A}^2$. *Let* $\mathbf{Q}$ *be a matrix* q_{ab}. *In the following we say* $\mathbf{P}$ *and* $\mathbf{Q}$ *are conjugate if there exists a non-zero complex vector* $(x_a)_{a\in\mathcal{A}}$ *such that* $q_{ab} = \frac{x_a}{x_b} p_{ab}$. *We say that such matrices are imaginary conjugate if* $|x_a| = 1$ *for all* $a \in \mathcal{A}$.

Observe that: (i) two conjugate matrices have the same eigenvalue set; (ii) if $\mathbf{u} = (u_a)_{a\in\mathcal{A}}$ is right eigenvector of $\mathbf{P}$, then $(x_a u_a)_{a\in\mathcal{A}}$ is right eigenvector of $\mathbf{Q}$. Similarly, if $(\zeta_a)_{a\in\mathcal{A}}$ is left eigenvector of $\mathbf{P}$, then $(\frac{1}{x_a}\zeta_a)_{a\in\mathcal{A}}$ is the left eigenvector of $\mathbf{Q}$.

The following lemma is essential and proved in [9] but we give an independent proof in the Appendix (see also [17]).

Lemma 5. *Let* $\mathbf{M} = [m_{ab}]_{(a,b)\in\mathcal{A}^2}$ *be a matrix such that* $m_{ab} \geq 0$. *We assume that 1 is the largest eigenvalue of* $\mathbf{M}$. *Let* $\mathbf{Q}$ *be a matrix with coefficients* $q_{ab} = e^{i\theta_{ab}} m_{ab}$ *where* θ_{ab} *is real. The matrix* $\mathbf{Q}$ *has eigenvalue 1 if and only if* $\mathbf{Q}$ *is imaginary conjugate to matrix* $\mathbf{M}$.

Corollary 1. *Let* $c \in \mathcal{A}$. *The matrix* $\mathbf{Q}$ *defined in Lemma 5 has eigenvalue 1 if and only if for all* $(a, b) \in \mathcal{A}^2$:

$$\frac{1}{2\pi} (\theta_{ab} + \theta_{ca} - \theta_{cb}) \in \mathbb{Z} \ . \tag{63}$$

Proof. If Q is conjugate to M, we should have a real vector $\theta_{a\in\mathcal{A}}$ such that $\forall (a, b) \in \mathcal{A}^2 \ \theta_{ab} = \theta_a - \theta_b$. Then $e^{i(\theta_a - \theta_b)} = \frac{e^{i\theta_{cb}}}{e^{i\theta_{ca}}}$, thus $e^{i(\theta_{cb} - \theta_{ca})} = e^{i\theta_{ab}}$. □

Lemma 6. *Let* $c \in \mathcal{A}$. *A tuple* (s_1, s_2) *belongs to* $\partial\mathcal{K}$ *iff for all* $(a, b) \in \mathcal{A}^2$ *we have*

$$\frac{\Im(s_1)}{2\pi} \log \frac{P_1(a|b)P_1(c|a)}{P_1(c|b)} - \frac{\Im(s_2)}{2\pi} \log \frac{P_2(a|b)P_2(c|a)}{P_2(c|b)} \in \mathbb{Z} \ . \tag{64}$$

Proof. Set $\mathbf{M} = \mathbf{P}(c_1, c_2)$ and $\mathbf{Q} = \mathbf{P}(s_1, s_2)$ for $(s_1, s_2) \in \partial\mathcal{K}$. Then, it follows directly from Corollary 1 with $e^{i\theta_{ab}} = (P_1(a|b))^{i\Im(s_1)}(P_2(a|b))^{-i\Im(s_2)}$. □

Furthermore, in the Appendix we prove the following important characterization of the set $\mathcal{K}$. We say that a curve is strictly concave (or strictly convex) if the is never linear, even locally.

Lemma 7. *If* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are not conjugate, then the set* $\mathcal{K}$ *is strictly concave.*

We summarize our knowledge about $\partial\mathcal{K}$.

Theorem 10. *There are three possible structures of* $\partial\mathcal{K}$:

- *the punctual case:* $\partial\mathcal{K} = \{(c_1, c_2)\}$, *this is the most typical case;*
- *the linear case: there exist a vector* $(x, y) \in \mathbb{R}^2$ *such that* $\partial\mathcal{K} = \{(c_1, c_2) + ik(x, y), k \in \mathbb{Z}\}$;
- *the lattice case: there exists two vectors* (x_1, y_1) *and* $(x_2, y_2) \in \mathbb{R}^2$ *which are not colinear such that* $\partial\mathcal{K} = \{(c_1, c_2) + ik_1(x_1, y_1) + ik_2(x_2, y_2), (k_1, k_2) \in \mathbb{Z}^2.\}$.

Proof. This follows from the fact that according to Lemma 6 if $(c_1, c_2) + (s_1, s_2) \in \partial\mathcal{K}$ then $\forall k \in \mathbb{Z}$ $(c_1, c_2) + k(s_1, s_2) \in \partial\mathcal{K}$. Furthermore if $(c_1, c_2) + (s_1', s_2') \in \partial\mathcal{K}$ then $(s_1, s_2) + a(s_1', s_2') \in \partial\mathcal{K}$. Thus $\mathcal{K}$ forms a lattice. In Lemma 6 this occurs when $\mathbf{P}_1$ and $\mathbf{P}_2$ are rationally related.

When both matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are logarithmically rationally related then we are in the lattice case, and the lattice is made of edges parallel to the axes. Anyhow the reverse is not necessarily true, although we don't know an explicit example of non logarithmically rationally related matrix which makes a pair of logarithmically commensurable matrices which would lead to edges non parallel to the axes.

When only one matrix is logarithmically rationally related, then we are in the linear case, and $\partial\mathcal{K}$ is a set of periodic points laying on one axis. It is nevertheless possible to have a linear case when none of the matrices is logarithmically rationally related, for example when $\mathbf{P}_1$ and $\mathbf{P}_2$ are of the form $\log^* \mathbf{P}_1 = 2\pi\mathbf{Q}_1 + \mathbf{M}$ and $\log^* \mathbf{P}_2 = -2\pi\mathbf{Q}_2 + \mathbf{M}$ where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have integer coefficients but $\mathbf{M}$ is not rationally related (in this case $x \log^* \mathbf{P}_1 + y \log^* \mathbf{P}_2$ integers would implies $x = y$. □

Now we establish some properties of the eigenvalue $\lambda(s_1, s_2)$ of $\mathbf{P}(s_1, s_2)$.

Lemma 8. *For all s_2 such that $\Re(s_2) = c_2$, assume $\nexists s_1 : (s_1, s_2) \in \partial\mathcal{K}$ then $\lambda(s_1, s_2) = 1 \Rightarrow \Re(s_1) < c_1$.*

Proof. Notice that $\Re(s_1) = c_1$ is not possible by construction since it would imply that $(s_1, s_2) \in \partial\mathcal{K}$. Let's consider the hypothesis $\Re(s_1) > c_1$. But we have $|\lambda(s_1, s_2)| \leq \lambda(\Re(s_1), \Re(s_2)$. Since $\Re(s_1) > c_1$, each non zero coefficient of $\mathbf{P}(\Re(s_1), \Re(s_2))$ are strictly smaller than the corresponding coefficients $\mathbf{P}(c_1, \Re(s_2))$ and therefore $\lambda(\Re(s_1), \Re(s_2)) < \lambda(c_1, \Re(s_2)) = 1$ which contradicts the hypothesis $\lambda(s_1, s_2) = 1$. □

Lemma 9. *We have $\lambda(c_1, c_2) > |\lambda_2(c_1, c_2)|$.*

Proof. It follows from Perron-Frobenius that the main eigenvalue is unique. □

Let $\mathcal{U}$ be a complex neighborhood of 0 such that $\forall s \in \mathcal{U}$: $|\lambda(c_2 + s)| > |\lambda_2(c_2 + s)|$. Therefore the function $\lambda(c_2 + s)$ is analytic. In the Appendix we prove the following lemma.

Lemma 10. *Let (x_k, y_k) be a sequence of complex numbers such that $\lim_{k\to\infty} \Re(x_k, y_k) = (c_1, c_2)$ and $|\lambda(x_k, y_k)| \to \lambda(c_1, c_2) = 1$. Then for all $(s_1, s_2) \in \mathcal{U}$ we have*

$$\forall j : \quad \lim_{k\to\infty} \frac{\lambda_j(x_k + s_1, y_k + s_2)}{\lambda(x_k + s_1, y_k + s_2)} = \frac{\lambda_j(c_1 + s_1, c_2 + s_2)}{\lambda(c_1 + s_1, c_2 + s_2)}, \tag{65}$$

and the function $\lambda(x_k + s_1, y_k + s_2)$ are all analytic and uniformly bounded functions on a complex neighborhood of $(0, 0)$ such that

$$\lim_{k\to\infty} \lambda(x_k + s_1, y_k + s_2) = \lambda(c_1 + s_1, c_2 + s_2) \tag{66}$$

$$\lim_{k\to\infty} \nabla\lambda(x_k + s_1, y_k + s_2) = \nabla\lambda(c_1 + s_1, c_2 + s_2) . \tag{67}$$

where ∇f is the gradient of f.

5 Proof of Theorem 3: Nilpotent Case

In this section we consider the case when the matrix $\mathbf{P}(s_1, s_2)$ is nilpotent, that is, there exists K such that $\mathbf{P}^K(s_1, s_2) = 0$ for all (s_1, s_2). We first provide a simple derivation, and then "recover" it through the Mellin approach.

Notice that for $z \in \mathbb{C}$

$$1 + z\langle \mathbf{1}_C | (\mathbf{I} - z\mathbf{P}(0, 0))^{-1}\mathbf{1}\rangle = 1 + \sum_{k\leq K} z^{k+1}\langle \mathbf{1}_C | \mathbf{P}^k(0, 0)\mathbf{1}\rangle$$

is the generating function that enumerates all the common words between the language of source 1 and the language of source 2, including the empty word. Let us call this set $\mathcal{W}$. Observe that $\langle \mathbf{1}_C | \mathbf{1}\rangle$ enumerate the word of length 1, and $|\mathcal{W}| = 1 + \langle \mathbf{1}_C | (\mathbf{I} - \mathbf{P}(0, 0))^{-1}\mathbf{1}\rangle$ is the total number of such common words. Notice that such words are all of length smaller than K. Since the Markov source are stationary we also notice that $\boldsymbol{\pi}(0, 0) = \mathbf{1}_C$.

The quantity $J_{n,m}$ converges to

$$1 + \langle \mathbf{1}_C | (\mathbf{I} - \mathbf{P}(0, 0))^{-1}\mathbf{1}\rangle$$

when $n, m \to \infty$ because all words in $\mathcal{W}$ will appear in both string almost surely. Indeed each word in $w \in \mathcal{W}$ may not appear in one string with exponentially small probability.

For similar reasons $C_{n,m}$ will converge to

$$1 + \langle \mathbf{1}_C | (\mathbf{I} - \mathbf{P}(0, 0))^{-1}\mathbf{1}\rangle$$

exponentially fast, because any word $w \in \mathcal{W}$ may be prefix to none of n independent strings with a probability decaying exponentially fast to 0.

Interestingly enough we can find partially this result via the reverse Mellin transform (62). Partially because the error term is $O(n^{-M})$ for all $M > 0$. Let

$$\begin{aligned} D(s_1, s_2) &= \langle \boldsymbol{\pi}(s_1, s_2) | \widetilde{\mathbf{C}}(s_1, s_2) + s_1 \widetilde{\mathbf{C}}(s_1, -1) \\ &\quad + s_1 \widehat{\mathbf{C}}(-1, s_2) + s_1 s_2 \widehat{\mathbf{C}}(-1, -1) \rangle. \end{aligned}$$

We notice that $D(s_1, s_2)$ is never singular and furthermore for all s $D(s, -1) = D(-1, s) = 0$. Let

$$D_n = \frac{1}{(2i\pi)^2} \int_{\rho_1} \int_{\rho_2} \Gamma(s_1)\Gamma(s_2) D(s_1, s_2) n^{-s_1 - s_2}.$$

Thus by (62) we find $C(n, n) = (1 - e^{-n})^2 + D_n$. Let M be an arbitrary non negative (large) number. By moving the integration path for s_2 from $\Re(s_2) = \rho_1$ to $\Re(s_2) = M$ we only met the poles of $\Gamma(s_2)$ on $s_2 = -1$ with residues

$$\frac{1}{2i\pi} \int_{\rho_1} \Gamma(s_1) D(s_1, -1) n^{1 - s_1} ds_1$$

and

$$-\frac{1}{2i\pi} \int_{\rho_1} \Gamma(s_1) D(s_1, 0) n^{-s_1} ds_1.$$

The first residues is null since $D(s_1, -1) = 0$, thus

$$\begin{aligned} D_n &= -\frac{1}{2i\pi} \int_{\rho_1} \Gamma(s_1) D(s_1, 0) n^{-s_1} ds_1 \\ &\quad + \frac{1}{(2i\pi)^2} \int_{\rho_1} \int_{M} \Gamma(s_1)\Gamma(s_2) n^{-s_1 - s_2} \end{aligned}$$

where the second term in the right-hand side is $O(n^{-M-\rho_1})$. The integration path $-\frac{1}{2i\pi} \int_{\rho_1} \Gamma(s_1) D(s_1, 0) n^{-s_1} ds_1$ can also be moved on $\Re(s_1) = M$, the residues on $s_1 = -1$ is $D(-1, 0)n$, which is null, and on $s_1 = 0$ is equal to $D(0, 0)$. Thus

$$D_n = D(0, 0) - \frac{1}{2i\pi} \int_{M} \Gamma(s_1) D(s_1, 0) n^{-s_1} + O(n^{-M-\rho_1}). \tag{68}$$

Since $\frac{1}{2i\pi} \int_M \Gamma(s_1) D(s_1, 0) n^{-s_1} = O(n^{-M})$ and that $D(0, 0) = \langle \mathbf{1}_C | (\mathbf{I} - \mathbf{P}(0, 0))^{-1} \mathbf{1} \rangle$, this concludes the proof.

6 Special Case: Proofs of Theorems 4 – 5

To simplify our presentation we will first assume that

$$\mathbf{P}_1 = \frac{1}{|\mathcal{A}|} \mathbf{1} \otimes \mathbf{1},$$

i.e. the first source is uniform and memoryless. We will see in the next section how to translate these results into the general case.

In this case, we have

$$\mathbf{P}(s_1, s_2) = |\mathcal{A}|^{s_1} \mathbf{P}(s_2) \tag{69}$$

with $\mathbf{P}(s) = \mathbf{P}(0, s)$. We also write $\pi(a) = \pi_2(a)$ and $\boldsymbol{\pi}(s) = \boldsymbol{\pi}(0, s)$, thus

$$\boldsymbol{\pi}(s_1, s_2) = |\mathcal{A}|^{s_1} \boldsymbol{\pi}(s_2) \ . \tag{70}$$

Let $\lambda(s_1, s_2)$ be the main (largest) eigenvalue of $\mathbf{P}(s_1, s_2)$. We have

$$\lambda(s_1, s_2) = |\mathcal{A}|^{s_1} \lambda(s_2) \tag{71}$$

where $\lambda(s)$ is the main eigenvalue of matrix $\mathbf{P}(s)$. We also define $\mathbf{u}(s)$ as the right eigenvector of $\mathbf{P}(s)$ and $\boldsymbol{\zeta}(s)$ as the left eigenvector provided $\langle \boldsymbol{\zeta}(s) | \mathbf{u}(s) \rangle = 1$.

We first present some simple results regarding $\lambda(s)$ and $L(s) = \log_{|\mathcal{A}|} \lambda(s)$.

Lemma 11. *The function $L(s)$ is convex when s is real.*

Proof. The function $(-L(s), s)$ describes the set $\mathcal{K}$ which is known to be a concave curve by Lemma 1. Notice that the proof will also be valid for the general case. $\square$

The proof of the following lemma is left for the reader.

Lemma 12. *We have the following identities:*

$$\begin{array}{rcl} \lambda(s) & = & \langle \boldsymbol{\zeta}(s)|\mathbf{P}(s)\mathbf{u}(s)\rangle = \sum_{a,b}\zeta_a(s)u_b(s)P(a|b)^{-s}, \\ \lambda'(s) & = & \langle \boldsymbol{\zeta}(s)|\mathbf{P}'(s)\mathbf{u}(s)\rangle = \sum_{a,b}\zeta_a(s)u_b(s)P(a|b)^{-s}(-\log P(a|b)), \\ \lambda''(s) & = & \langle \boldsymbol{\zeta}(s)|\mathbf{P}''(s)\mathbf{u}(s)\rangle = \sum_{a,b}\zeta_a(s)u_b(s)P(a|b)^{-s}(\log P(a|b))^2. \end{array} \quad (72)$$

Finally, to compute some of the constants in Theorems 4 – 5, we need to computer $L''(s)$. To do so, let $x_{a,b} = \frac{1}{\lambda(s)}\zeta_a(s)u_b(s)P(a|b)^{-s}$. Clearly, by Lemma 12 we have $\sum_{a,b} x_{a,b} = 1$ and

$$L''(s) = \sum_{a,b} x_{a,b}(\log P(a|b))^2 - \left(\sum_{a,b} x_{a,b}\log P(a|b)\right)^2. \quad (73)$$

Now we are ready to derive our results presented in Theorems 4 – 5. The starting point is the Mellin transform $C^*(s_1,s_2)$ shown in (61) with $\widetilde{\mathbf{C}}^*(s_1,s_2)$ presented in (57). To recover $C_{n,n}$ we first need to find the inverse Mellin transform of (61). For $-2<\rho<-1$ we have

$$\begin{array}{rl} & \mathbf{C}(z,z) - \mathbf{D}(z,z) = \\ & \frac{1}{(2i\pi)^2}\int_{\Re(s_1)=\Re(s_2)=\rho} \widetilde{\mathbf{C}}^*(s_1,s_2) z^{-s_1-s_2} ds_1 ds_2 \\ = & \frac{1}{(2i\pi)^2}\int_{\Re(s_1)=\Re(s_2)=\rho} \Gamma(s_1)\Gamma(s_2)\Big(\widehat{\mathbf{C}}(s_1,s_2) + s_2\widehat{\mathbf{C}}(s_1,-1) \\ & +s_1\widehat{\mathbf{C}}(-1,s_2) + s_1 s_2\widehat{\mathbf{C}}(-1,-1)\Big) z^{-s_1-s_2} ds_1 ds_2, \end{array}$$

where $\widehat{\mathbf{C}}(s_1,s_2) = (\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1}\,\mathbf{1}$ and

$$\widetilde{\mathbf{C}}(s_1,s_2) = \langle \boldsymbol{\pi}(s_1,s_2)\,(\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1}\,\mathbf{1}\rangle.$$

Since $\mathbf{C}(z,z) = \mathbf{D}(z,z) + O(z^{-M})$ for any $M>0$ when $\Re(z)\to\infty$ we find

$$\begin{array}{rcl} C(z,z) & - & 1 + O(z^{-M}) = \frac{1}{(2i\pi)^2}\int_{\Re(s_1)=\Re(s_2)=\rho} \Gamma(s_1)\Gamma(s_2) \\ & \times & \Big(\widehat{\mathbf{C}}(s_1,s_2) + s_2\widehat{\mathbf{C}}(s_1,-1)\ \ s_1\widehat{\mathbf{C}}(-1,s_2) + s_1 s_2\widehat{\mathbf{C}}(-1,-1)\Big) z^{-s_1-s_2} ds_1 ds_2. \end{array} \quad (74)$$

To analyze it asymptotically, we investigate the set of singularities of $(\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1}$ in $\widehat{\mathbf{C}}(s_1,s_2)$. Recall that $\mathcal{K}$ is the set of complex numbers (s_1,s_2) such that $\mathbf{I}-\mathbf{P}(s_1,s_2)$ is degenerate, *i.e.* $(\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1}$ is singular.

Let $\lambda_1(s), \lambda_2(s)\ldots,\lambda_{|\mathcal{A}|}(s))$ be the eigenvalues of $\mathbf{P}(s)$ in the non-increasing order (e.g., $\lambda(s) := \lambda_1(s))$ while $\mathbf{u}_i(s)$ and $\boldsymbol{\zeta}_i(s))$ are respectively the right and the left eigenvectors of $\mathbf{P}(s)$ associated with $\lambda_i(s)$ subject to $\langle\boldsymbol{\zeta}_i(s)|\mathbf{u}_i(s)\rangle = 1$. By the spectral representation of matrices [23], we have

$$(\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1} = \sum_{i=1}^{|\mathcal{A}|}\frac{1}{1-|\mathcal{A}|^{s_1}\lambda_i(s_2)}\mathbf{u}_i(s_2)\otimes\boldsymbol{\zeta}_i(s_2) \quad (75)$$

where $\otimes$ denotes the tensor product. Observe that $(\mathbf{I}-\mathbf{P}(s_1,s_2))^{-1}$ cease to exist at (s_1,s_2) satisfying $|\mathcal{A}|^{s_1}\lambda_i(s_2) = 1$, that is, for $s_1 := L_{i,k}(s_2)$ where

$$L_{i,k}(s_2) = \frac{1}{-\log|\mathcal{A}|}(\log\lambda_i(s_2) + 2ik\pi).$$

The eigenvalues $\lambda_i(s)$ are individually analytic functions of s in any complex neighborhood where the order of the eigenvalues modulus does not change (*i.e.* $|\lambda_{i-1}(s)| > |\lambda_i(s)| > |\lambda_{i+1}(s)|$ for all i). But any function of the form $\sum_i f(\lambda_i(s))$ is analytic even when the eigenvalue sequence is not strictly decreasing, as long as $f()$ is analytic. To simplify our analysis, we also postulate that none of the eigenvalue is identically equal to zero, that is, we assume $\log \lambda_i(s)$ exists except on a countable set $\mathcal{R} = \{s :\ \exists i : \lambda_i(s) = 0\}$. It should be pointed out that there are cases when some eigenvalues are identically equal to zero. For example, for memoryless sources we have for all $i \geq 2$: $\lambda_i(s) \equiv 0$ which we already discussed in [6, 10] so we will omit them here.

In order to evaluate the integral in (74) we first use (75) and then apply the residue theorem. To simplify, for $1 \leq j \leq |\mathcal{A}|$, let $f_j(s) = \langle \boldsymbol{\pi}(s)|\mathbf{u}_i(s)\rangle$ and $g_j(s) = \langle \boldsymbol{\zeta}_i(s)|\mathbf{1}\rangle$. Define (here we set $s := s_2$)

$$I(z,\rho) = \tag{76}$$

$$\frac{1}{2i\pi}\int_{\Re(s)=\rho}\sum_{k\in\mathbb{Z}}\sum_{j=1}^{|\mathcal{A}|}\frac{f_j(s)g_j(s)\Gamma(-L_{j,k}(s))\Gamma(s)}{\lambda_j(s)\log|\mathcal{A}|}z^{L_{j,k}(s)-s}ds$$

$$J(z,\rho) = \frac{1}{2i\pi}\int_{\Re(s)=\rho}\widehat{C}(0,s)\Gamma(s)z^{-s}ds. \tag{77}$$

Furthermore, let

$$H_j(s,z) = \sum_{k\in\mathbb{Z}}\frac{f_j(s)g_j(s_2)\Gamma(-L_{j,k}(s))}{\lambda_i(s)\log|\mathcal{A}|}z^{L_{j,k}(s)} \tag{78}$$

thus

$$I(z,\rho) = \frac{1}{2i\pi}\sum_j\int_{\Re(s)=\rho}H_j(s,z)\Gamma(s)z^{-s}ds.$$

The next lemma is crucial for the asymptotic evaluation of $C(z,z)$ which by depoissonization lead to asymptotics of $C_{n,n}$ and ultimately $J_{n,n}$.

Lemma 13. *For any $M > 0$ and for some $\rho > -1$, we have*

$$C(z,z) = 1 + I(z,\rho) - J(z,\rho) + O(z^{-M}) \tag{79}$$

for $z \to \infty$.

Proof. In the inverse Mellin expression we see that for $\Re(s_1) = \Re(s_2) = -1$ we have $|\mathbf{P}(s_1,s_2)| \leq \mathbf{P}(-1,-1)$ and $\mathbf{P}(-1,-1) \leq \mathbf{P}_1(-1)$ and $\mathbf{P}_2(-1)$. Since the matrix $\mathbf{P}(s_1,s_2)$ is not nilpotent there exists $(a,b) \in \mathcal{A}^2$ such that $|\mathbf{P}(s_1,s_2)_{a,b}| < \mathbf{P}(a|b)$. Consequently, there exists k such that $|\mathbf{P}^k(s_1,s_2)|\mathbf{1} < \mathbf{1}$ or more precisely $|\lambda_1(s_1,s_2)| \leq 1-\epsilon'$ for some $\epsilon' > 0$. Thus there exists $\epsilon > 0$ such that for all s_1, s_2 $\Re(s_1) > -1+\epsilon$ and $\Re(s_2) > -1+\epsilon$ implies that $\mathbf{I} - \mathbf{P}(s_1,s_2)$ is not degenerate.

To evaluate the inverse Mellin transform we apply standard approach by moving the line of integration to "catch up" relevant singularities, however, in our case there some complications. We move the integration path by increasing ρ. This does not change the value of $I(z,\rho)$ and $J(z,\rho)$ as long as the functions in the integral paths are analytic and not singular. When the path encounter a singularity we will use the residue theorem. But we may have a problem when any of the functions λ_j ceases to be analytic. However, we shall see that when we sum all the terms of the integrand of $I(z,\rho)$ we obtain an analytic function derived from $\langle \boldsymbol{\pi}(s_1,s_2)(\mathbf{I} - \mathbf{P}(s_1,s_2))^{-1}\mathbf{1}\rangle$. Indeed we have the (somewhat complicated) identity

$$I(z,\rho) = \sum_{k\in\mathbb{Z}}\int_{\Re(s)=\rho}\langle \boldsymbol{\pi}(s)|(\mathbf{P}(s))^{-1}\exp\left(-\frac{\log z}{\log|\mathcal{A}|}(\log\mathbf{P}(s)+2ik\pi\mathbf{I})\right)\Gamma\left(-\frac{1}{\log|\mathcal{A}|}(\log\mathbf{P}(s)+2ik\pi\mathbf{I})\right)\mathbf{1}\rangle ds\ , \tag{80}$$

knowing that any analytical function $f(.)$ can be applied to matrix $\mathbf{P}(s)$ as long its eigenvalues do not correspond to a singularity of the function $f(.)$. Therefore the only singularities that we meet when we move the integration line of $I(z,\rho)$ are the elements of $\mathcal{R} = \{s :\ \lambda_i(s) = 0, \text{for some } i\}$.

If $\theta \in \mathcal{R}$, is one of these singularity, thus we have $\lambda_i(\theta) = 0$, then the function $L_{i,k}(s) = \frac{1}{-\log|\mathcal{A}|}(\log\lambda_i(s) + 2ik\pi)$ is meromorphic around θ. However if θ is a simple pole of $\lambda_i(s)$, then moving around θ would be equivalent to add 1 to the integer k: $\log \lambda_i(s) \to \log \lambda_i(s) + 2i\pi$. If the root is of multiplicity ℓ it is equivalent to add ℓ to the integer k. In any case the function $H_i(s,z)$ being invariant when ℓ is added to k, turns out to be fully analytic around θ, and the integration path in $I(z,\rho)$ can be moved over θ.

However, the function $\lambda_i(s)$ is a non polar singular on $s = \theta$, hence there will be a contribution coming from the integration of $H_i(s,z)\Gamma(s)z^{-s}$ on an arbitrary small loop around θ. Since $\Re(L_{i,k}(s)) \to -\infty$ when $s \to \theta$, having $\Re(L_{i,k}(s)) < -M$ will guarantee that the contribution is in $O(z^{1-M})$ and can be included in the error term.

Moving the integration path from $\Re(s_1) = \Re(s_2) = \rho$ to $\Re(s_1) = \Re(s_2) = -1 + \epsilon$ will only hit the poles of $\Gamma(s_1)\Gamma(s_2)$ at $s_1 = -1$ and $s_2 = -1$. By construction of the function $\widehat{\mathbf{C}}(s_1, s_2) + s_1\widehat{\mathbf{C}}(-1, s_2) + s_2\widehat{\mathbf{C}}(s_1, -1) + s_1 s_2 \widehat{\mathbf{C}}(-1,-1)$, the residues at these points are zero. Therefore the expression

$$\begin{aligned} C(z,z) \quad - \quad & 1 + O(z^{-M}) = \frac{1}{(2i\pi)^2}\int_{\Re(s_1)=\Re(s_2)=\rho} \Gamma(s_1)\Gamma(s_2) \\ & \times \Big(\widehat{\mathbf{C}}(s_1,s_2) + s_2\widehat{\mathbf{C}}(s_1,-1) \\ & s_1\widehat{\mathbf{C}}(-1,s_2) + s_1 s_2\widehat{\mathbf{C}}(-1,-1)\Big)\, z^{-s_1-s_2} ds_1 ds_2 \end{aligned}$$

still holds for $\rho = -1 + \epsilon$.

Now we take the integration contour for s_1 and we move it from $\Re(s_1) = \rho$ to $\Re(s_1) = M - \rho$. By doing so we encounter many poles:
(i) The poles of $(\mathbf{I} - \mathbf{P}(s_1, s_2))^{-1}$ at $s_1 = L_{j,k}(s_2)$. The residues is exactly the expression $I(z,\rho)$.
(ii) The poles of $\widehat{C}(s_1, s_2)\Gamma(s_1)$ at $s_1 = 0$ which has residues $-J(z,\rho)$.
(iii) The double pole of $\widehat{C}(s_1,-1)\Gamma(s_1)z^{-s_1}$ at $s_1 = 0$ since $(\mathbf{I} - \mathbf{P}(s_1,s_2))^{-1}$ is singular at $(s_1,s_2) = (0,-1)$ because $\lambda_1(0,-1) = 1$. It leads to the residue

$$-\frac{1}{2i\pi}\int_{\Re(s)=\rho} s\Gamma(s)(a\log z + b)z^{-s} ds \tag{81}$$

for some real number a and b coming from the derivative of $f_1(s)$ and $g_1(s)$ at $s = 0$. But when one moves the integration path of (81) to $\Re(s) = M$ the function $s\Gamma(s)(a\log z + b)$ has no singularity since $s\Gamma(s)$ is not singular on the interval $]-1, +\infty[$, and thus the integration on $\Re(s) = M$ is $O(z^{-M})$, which can be included in the error term. □

In the following we denote $L(s) = L_{1,0}(s)$. The rule of the game is that we move the integration abscissa of $I(z,\rho)$ and $J(z,\rho)$ to the left (*i.e.* to larger values) on the value c_2 which minimizes the argument $L(s) - s$. Moving the integration path one meets some poles of $\Gamma(-L_j(s))$ when $L_j(s) = 0$. In fact when the matrices are strictly non negative, this case only applies to $j \geq 2$. It turns out that when s is a pole for $\Gamma(-L_j(s))$ then it is at the same time a pole of $\widehat{C}(0,s)$. The residues of $I(z,\rho)$ and $J(z,\rho)$ when passes over such value are the same and cancel. Therefore $C(z,z) = 1 + I(z,\rho) - J(z,\rho) + O(z^{-M})$ for all values of $\rho < 0$.

6.1 Proof of Theorems 4 and 5

Now we are going to prove Theorems 4 and 5 corresponding to the case where quantities c_1 and c_2 are both in the interval $[-1,0]$ in the case where one source is uniform memoryless. In this case, the main contribution to $C(z,z)$ doesn't come from the poles, as in the previous section, but rather from the saddle point of $z^{L_{1,0}-s}$ (in fact, infinitely many saddle points).

We start with reviewing some properties of the kernel $\mathcal{K}$ and the main eigenvalue. Recall that $\partial\mathcal{K}$ is the set of complex tuples (s_1,s_2) satisfying $|\mathcal{A}|^{s_1}\lambda(s_2) = 1$ such that $\Re(s_1) = c_1$ and $\Re(s_2) = c_2$. Its structure is crucial for our asymptotic analysis.

From the general Theorem 10 we deduce that only two cases are possible when one source, say source 1, is uniform (since $\mathbf{P}_1$ is logarithmically rationally related):

- the lattice case when $\mathbf{P}_2$ is also logarithmically rationally related, we call this case the *rational case*;

- the linear case when $\mathbf{P}_2$ is not logarithmically rationally related, we call this case the *irrational case*;

Now we focus on proving in the next four lemmas that the main eigenvalue is well separated.

Lemma 14. *Let t_2 be a real number. We have the equivalence*

$$\nexists (s_1, s_2) \in \partial\mathcal{K} \quad \Im(s_2) = t_2 \iff |\lambda(c_2 + it_2)| < \lambda(c_2).$$

Proof. Let $s_2 = c_2 + it_2$. By the Perron-Frobenius, we have $|\lambda(s_2)| \leq \lambda(c_2)$ since $\Re(s_2) = c_2$ and $|\mathbf{P}(s_2)| = \mathbf{P}(c_2)$ (by taking the modulus element-wise). If $|\lambda(s_2)| = \lambda(c_2)$, then there will be t_1 such that $|\mathcal{A}|^{it_1}\lambda(s_2) = \lambda(c_2)$, and therefore $(c_1 + it_1, s_2) \in \partial\mathcal{K}$. □

Lemma 15. *We have a non zero spectral gap, that is, $\lambda(c_2) > \lambda_2(c_2)$.*

Proof. It follows from Perron-Frobenius that the main eigenvalue is unique. □

Let $\mathcal{U}$ be a complex neighborhood of 0 such that $\forall s \in \mathcal{U}$: $|\lambda(c_2 + s)| > |\lambda_2(c_2 + s)|$. Therefore the function $\lambda(c_2 + s)$ is analytic.

Lemma 16. *Let s_k be a sequence such that $\Re(s_k) = c_2$ and $|\lambda(s_k)| \to \lambda(c_2)$. Then for all $s \in \mathcal{U}$ we have*

$$\lim_{k\to\infty} L(s_k + s) - L(s_k) = L(c_2 + s) - L(c_2) \tag{82}$$

$$\lim_{k\to\infty} L'(s_k + s) = L'(c_2 + s) \ . \tag{83}$$

The convergence also holds for any derivative of function $L'(s)$, and the function $\lambda(s_k + s)$ is analytic and uniformly bounded on a complex neighborhood of 0.

Proof. It turns out that $\lim_k |\lambda(c_1, s_k)| = |\mathcal{A}|^{c_1}\lambda(c_2) = 1$. There exists x_k such that $\Re(x_k) = c_1$ and $\lambda(x_k, s_k) = |\lambda(c_1, s_k)|$. Hence Lemma 10 applies. Thus for any complex number s

$$\begin{aligned}\forall j: \quad \lim_{k\to\infty} \frac{\lambda_j(s_k + s)}{\lambda(s_k + s)} &= \lim_{k\to\infty} \frac{\lambda_j(x_k, s_k + s)}{\lambda(x_k, s_k + s)} \\ &= \frac{\lambda_j(c_1, c_2 + s)}{\lambda(c_1, c_2 + s)} = \frac{\lambda_j(c_2 + s)}{\lambda(c_2 + s)}\end{aligned} \tag{84}$$

Since $\frac{|\lambda_2(c_2)|}{\lambda(c_2)} < 1$, there exists $\mathcal{U}$ such that $\forall s \in \mathcal{U}$: $\left|\frac{\lambda_j(c_2+s)}{\lambda(c_2+s)}\right| < 1$ thus $\lambda(c_2 + s)$ is analytic because it never cross the value of another eigenvalue and so is $\lambda(s_k + s)$.

Hence, the logarithm of the eigenvalue, $L(s_k + s) - L(s_k)$ converges to $L(c_2 + s) - L(c_2)$. The property $|\lambda(c_1 + s, c_2 + s_2)| > \lambda_2(c_2 + s)$ for all $s \in \mathcal{U}$ implies the analyticity of $L(c_2 + s)$, and therefore $L'(c_2 + s)$. □

In passing, we have $L'(s_k) \to 1$ and $L''(s_k) \to \alpha_2$.

Finally, we prove that the main eigenvalue dominates all other eigenvalues in a complex neighborhood of c_2.

Lemma 17. *There exists $\epsilon > 0$ such that for all $i \neq 1$ and for all s such that $\Re(s) = c_2$:*

$$|\lambda_i(s)| < \lambda(c_2) - \epsilon \ . \tag{85}$$

Proof. This is a consequence of previous lemmas. Suppose that there exists s_k such that $|\lambda_2(s_k)| \to \lambda(c_2)$. This implies that $|\lambda(s_k)| \to \lambda(c_2)$, but by previous lemma $|\lambda_2(s_k)| \to \lambda_2(c_2) = \lambda(c_2) - \epsilon$. □

Now we are in the position to evaluate the integral of $C(z, z)$ by the saddle point methods. Recall that for all $M > 0$ we have $C(z, z) = I(z, c_2) - J(z, c_2) + 1 + O(z^{1-M})$ where $I(z, \rho)$ and $J(z, \rho)$ are given by (76). We already prove that $J(z, c_2) = O(z^{c_2}) = O(z^{\kappa-\epsilon})$ for some $\epsilon > 0$ (in fact $\epsilon = L(c_2) > 0$). We reinforce it in the next lemma.

Lemma 18. *There exists $\epsilon > 0$ such that*

$$C(z,z) = 1 + \frac{1}{2i\pi} \int_{\Re(s)=c_2} H_1(s,z)\Gamma(s) z^{-s} ds + O(z^{\kappa-\epsilon}) \tag{86}$$

where $\kappa = -c_1 - c_2$ and we recall that

$$H_1(z,s) = \sum_{k\in\mathbb{Z}} \frac{f_1(s)g_1(s)\Gamma(-L_{1,k}(s))}{\lambda(s)\log|\mathcal{A}|}.$$

Proof. By Lemma 17 for all $j > 1$ we have $\log_{|\mathcal{A}|}|\lambda_j(s)| < \log_{|\mathcal{A}|}\lambda(c_2) - \epsilon$ for some $\epsilon > 0$, thus the contribution of $\int_{\Re(s)=c_2} H_j(s,z)\Gamma(s)z^{-s}ds$ is of order

$$\int_{\Re(s)=c_2} |\Gamma(s)| z^{\Re(L_j(s)-s)} ds = O\left(z^{L(c_2)-c_2-\epsilon}\right) = O\left(z^{\kappa-\epsilon}\right),$$

as desired. $\square$

Rational Case. We assume now that the matrix $\log^*(\frac{1}{P(c|c)}\mathbf{P})$ is *rationally balanced*. The matrix $\mathbf{P}(s+2i\pi\nu)$ is then imaginary conjugate with the matrix $P(c|c)^{2i\pi\nu}\mathbf{P}(s)$ and $L(s+2i\pi\nu) = L(s) + 2i\pi\nu\log P(c|c)$. Thus $\Re(L(c_2+it))$ is periodic in t with period $2\pi\nu$. Furthermore, $L'(s)$ is also periodic with period $2\pi\nu$. Thus, $s_\ell = c_2 + 2i\pi\ell\nu$ for $\ell \in \mathbb{Z}$ are saddle points of $z^{L(s)-s}$.

We concentrate now on the term $k=0$ in $H_1(s,z)\Gamma(s)z^{-s}$ in (86). Define

$$b_2(s) = \frac{d^2}{ds^2}\log\left(\frac{f_1(s)g_1(s)}{\lambda_1(s)}\Gamma(-L(s))\Gamma(s)\right). \tag{87}$$

Notice that $b_2(s) = \beta_2(-L(s),s)$ mentioned in Theorem 8. Since the function

$$\log\left(\frac{f(s)g(s)}{\lambda(s)}\Gamma(-L(s))\Gamma(s)\right)$$

has bounded variations, we have the classic saddle point result [4, 23]

$$\begin{aligned} \frac{1}{2i\pi}\int_{\Re(s)=c_2} \frac{f(s)g(s)}{\lambda(s)}\Gamma(-L(s))\Gamma(s)z^{L(s)-s}ds = \\ = \sum_\ell \frac{f(s_\ell)g(s_\ell)}{\lambda(s_\ell)}\Gamma(-L(s_\ell))\Gamma(s_\ell) \\ \times \frac{z^{L(s_\ell)-s_\ell}}{\sqrt{2\pi(\alpha_2\log z + b_2(s_\ell))}}(1+o(1)). \end{aligned} \tag{88}$$

Notice that $\Re(L(s_\ell)-s_\ell) = \kappa$. When adding the contribution from the $L(s) + \frac{2ik\pi}{\log|\mathcal{A}|}$ we obtain the expression for $Q(\log z)$ with $\partial\mathcal{K} = \{(-L(s_\ell) - \frac{2ik\pi}{\log|\mathcal{A}|}, s_\ell), (k,\ell) \in \mathbb{Z}^2\}$. The double periodicity comes from the fact that $\sqrt{x}Q(x) = \sum_{k,\ell} q_{k,\ell}e^{i(k\alpha+\ell\beta)x} + o(1)$ when $x \to \infty$ for some *incommensurable*[3] pair of real numbers (α,β) and complex numbers $\{q_{k,\ell}\}_{(k,\ell)\in\mathbb{Z}^2}$.

Irrational Case. We now turn to the irrational case. Let $A > 0$ be a number such that for all $|s| \le A$ we have $|\lambda(c_2+s)| > |\lambda_2(c_2+s)|$; thus $L(c_2+s)$ is analytic. We assume that $c_2 < 0$ is the only saddle point on $\Re(s) = c_2$ for $|\Im(s)| \le A$. There also exists $\alpha_3 > 0$ such that

$$|t| \le A \Rightarrow \quad \Re(L(c_2+it) - L(c_2)) \le -\alpha_3 t^2\ . \tag{89}$$

[3] recall that a pair of numbers (α,β) is *commensurable* if there exists a real number ν such that the vector $(\nu\alpha,\nu\beta) \in \mathbb{Z}^2$; otherwise the pair is *incommensurable*.

From the previous analysis we know that

$$\begin{aligned}\frac{1}{2i\pi}\sum_{k\in\mathbb{Z}} & \int_{\Re(s)=c_2,|\Im(s)|\le A}\frac{f(s)g(s)}{\lambda(s)\log|\mathcal{A}|} \\ & \times\Gamma\left(-L(s)-\frac{2ik\pi}{\log|\mathcal{A}|}\right) \\ & \times\Gamma(s)z^{L(s)-s+2ik\pi/\log|\mathcal{A}|}ds \\ = & \ Q(\log z)(1+o(1)). \end{aligned}\tag{90}$$

Assume now (89) and define

$$\xi(s)=\sum_{k\in\mathbb{Z}}\left|\Gamma\left(s-\frac{2ik\pi}{\log|\mathcal{A}|}\right)\right| \ .\tag{91}$$

The function $\xi(s)$ is continuous and bounded as long as $\Re(s)$ is bounded. Our aim is to prove that

$$\begin{aligned}\frac{1}{2i\pi} & \int_{\Re(s)=c_2,|\Im(s)|>A}\left|\frac{f(s)g(s)}{\lambda(s)}\right|\xi(-L(s)) \\ & \times|\Gamma(s)|z^{\Re(L(s))-c_2}ds=o(\frac{z^\kappa}{\sqrt{\log z}}) \ ,\end{aligned}\tag{92}$$

which will complete the proof of Theorem 4.

We know that $|f(s)g(s)|\le f(c_2)g(c_2)$. In addition, we know that for $\Re(s)=c_2$ we have $\Re(L(s))<L(c_2)$ as long as $\Im(s)\neq 0$. We also have $|\lambda(s)|>\epsilon'$ for some $\epsilon'>0$ since the matrix $\mathbf{P}(s)$ stays away from the null matrix. Therefore, we need to estimate

$$\int_{\Re(s)=c_2,|\Im(s)|>A}|\Gamma(s)|z^{\Re(L(s))-c_2}ds \ .\tag{93}$$

For any $\epsilon>0$, the portion of the line $\Re(s)=c_2$, where $\Re(L(s))<L(c_2)-\epsilon$, contributes $z^{\kappa-\epsilon}$ to $C(z,z)$. Our attention must turn to the values of s on this line such that $\Re(L(s))$ is arbitrary close to $L(c_2)$. In particular, we are interested in the local maxima of $\Re(L(s))$ that are arbitrary close to $L(c_2)$. Indeed, these local maxima play a role in the saddle point method.

Let us consider the sequence of those maxima denoted by s_ℓ for $\ell\in\mathbb{N}$ such that $\Re(L(s_\ell))\to L(c_2)$. By Lemma 10 we know that for all real t $L(s_\ell+it)-L(s_\ell)\to L(c_2+it)-L(c_2)$ and that $L'(s_\ell+it)\to L'(c_2+it)$. Therefore for all real t such $|t|\le A$

$$\limsup_{\ell\to\infty}\left(\Re(L(s_\ell+it))-\Re(L(s_\ell))\right)\le-\alpha_3t^2.\tag{94}$$

We define $I(A)$ to be the set of complex numbers s such that $\Re(s)=c_2$ and $\min_\ell\{|s-s_\ell|\}>A$.

Lemma 19. *There exists ϵ such that for all $s\in I(A)$: $\Re(L(s))<L(c_2)-\epsilon$.*

Proof. Assume $s\in I(A)$. Since s is not a local maxima, we study the variation of $\Re(L(s))$ around the local maxima s_ℓ. Without loss of generality we assume that $s_\ell-A$ is between s and s_ℓ, thus $\Re(L(s_\ell-A))>\Re(s)$. Since $\limsup\Re(L(s_\ell-A))<L(c_2)-\alpha_3A^2<L(c_2)-\epsilon$ the lemma is proved. □

In view of the above, we conclude that

$$\begin{aligned}&\int_{\Re(s)=c_2,|\Im(s)|>A}|\Gamma(s)|z^{\Re(L(s))-c_2}ds\le \\ \sum_\ell & \int_{|t|\le A}|\Gamma(s_\ell+it)|z^{\Re(L(s_\ell+it))-c_2}dt+O(z^{\kappa-\epsilon}) \ .\end{aligned}\tag{95}$$

By virtue of the properties of function $\Gamma(s)$ on the imaginary lines, there exists a real $B>0$ such that $\forall s$:

$$\Re(s)=c_2\Rightarrow\max_{|t|\le A}\{|\Gamma(s+it)|\}\le B|\Gamma(s)| \ .\tag{96}$$

Therefore, our analysis can be limited to

$$\sum_{\ell} \int_{|t|\leq A} |\Gamma(s_\ell)| z^{\Re(L(s_\ell+it))-c_2} dt \ . \tag{97}$$

Finally, we establish a separation result.

Lemma 20. *For ℓ tending to infinity, the s_ℓ are separated by a distance at least equal to A.*

Proof. First, let us assume that $\ell, \ell' \to \infty$ and $|s_\ell - s_{\ell'}| \to 0$, then we have

$$L'(s_{\ell'}) = L'(s_\ell) + (s_{\ell'} - s_\ell)L''(s_\ell) + O(|s_\ell - s_{\ell'}|^2). \tag{98}$$

Since $L''(s_\ell) \to \alpha_2 \neq 0$, then we cannot have $L'(s_{\ell'}) = 1$, thus $s_{\ell'}$ cannot be a local maximum of $\Re(L(s))$. Second, if $\liminf |s_\ell - s_{\ell'}| > \epsilon$ for some $\epsilon > 0$ with $|s_\ell - s_{\ell'}| < A$, then using the inequality

$$\limsup \Re(L(s_{\ell'})) - \Re(s_\ell) \leq -\alpha_3 |s_\ell - s_{\ell'}|^2 < -\alpha_3 \epsilon^2 \tag{99}$$

we cannot have $\Re(L(s_{\ell'})) \to L(c_2)$. □

From the above we conclude

$$\begin{aligned} &\int_{|t|\leq A} z^{\Re(L(s_\ell+it))-c_2} dt = \\ &z^{\kappa} z^{\Re(s_\ell)-L(c_2)} \int_{|t|\leq A} z^{\Re(L(s_\ell+it))-\Re(L(s_\ell)} dt \ . \end{aligned} \tag{100}$$

In summary, the consequence of the previous lemma is that since $\limsup_{\ell\to\infty} \Re(L(s_\ell + it)) - \Re(L(s_\ell) \leq -\alpha_3 t^2$, we have [23]

$$\limsup_{\ell\to\infty} \int_{|t|\leq A} z^{\Re(L(s_\ell+it))-\Re(L(s_\ell)} dt \leq \frac{1}{\sqrt{\pi\alpha_3 \log z}}, \tag{101}$$

and the properties of function $\Gamma(s)$ is that $\sum_\ell |\Gamma(s_\ell)| < \infty$. Therefore,

$$\begin{aligned} \sum_{\ell} \quad & \int_{|t|\leq A} |\Gamma(s_\ell)| z^{\Re(L(s_\ell+it))-c_2} dt \\ = \quad & z^{\kappa} \sum_{\ell} |\Gamma(s_\ell)| z^{\Re(s_\ell)-L(c_2)} \\ & \times \int_{|t|\leq A} z^{\Re(L(s_\ell+it))-\Re(L(s_\ell)} dt \end{aligned} \tag{102}$$

since $\lim_{z\to\infty} z^{\Re(s_\ell)-L(c_2)} = 0$, by the dominating convergence theorem. We finally arrive at

$$\begin{aligned} &\sum_{\ell} |\Gamma(s_\ell)| z^{\Re(s_\ell)-L(c_2)} \int_{|t|\leq A} z^{\Re(L(s_\ell+it))-\Re(L(s_\ell)} dt \\ &= o\left(\frac{1}{\sqrt{\log z}}\right) \ . \end{aligned} \tag{103}$$

In fact, the saddle point expansion is extendible to any order of $\frac{1}{\sqrt{\log n}}$. This proves Theorems 4 and 5. In passing we observe that when $\mathbf{P}_1$ and $\mathbf{P}_2$ are logarithmically commensurable, the line $\Re(s_1) = c_1$ contains an infinite number of saddle points that contribute to the double periodic function $Q_2(\log n)$ (cf. [9] for more details).

7 General Case: Proofs of Theorem 6 – 9

We now look at the general case when $\mathbf{P}_1 \neq \mathbf{P}_2$. The main difficulty of the general case is when $\mathbf{P}_1$ and $\mathbf{P}_2$ have some zero coefficients, not at the same locations. For example $\mathbf{P}(-1,0)$ may differ from $\mathbf{P}_1$, and $\mathbf{P}(0,-1)$ may differ from $\mathbf{P}_2$ since $\mathbf{P}(s_1,s_2)$ retains only the coefficients that are both non zero in $\mathbf{P}_1$ and $\mathbf{P}_2$. For example, $\mathbf{P}(-1,0)$ and $\mathbf{P}(0,-1)$ may be conjugate while $\mathbf{P}_1$ and $\mathbf{P}_2$ are not. Indeed we can have $\mathbf{P}(-1,0) = \mathbf{P}(0,-1)$ even when $\mathbf{P}_1 \neq \mathbf{P}_2$.

In this section we first prove Theorem 6 which consider the case when $\mathbf{P}(-1,0)$ and $\mathbf{P}(0,-1)$ are conjugate. Then we present a detailed proof of Theorem 7, and finally we briefly discussed proofs of Theorems 8– 9.

Proof of Theorem 6. When $\mathbf{P}(-1,0)$ and $\mathbf{P}(0,-1)$ are conjugate, the situation is more similar to the case when $\mathbf{P}-1 = \mathbf{P}_2$ discussed in Section 4. In particular there is no saddle point, and the analysis reduces to computing some residues of poles.

To start, we notice that in this case there exists a vector of real numbers $(x_a)_{a\in\mathcal{A}}$ such that

$$P_1(a|b)P_2(a|b) > 0 \Rightarrow P_2(a|b)\frac{x_a}{x_b}P_1(a|b).$$

As a consequence we have the same spectrum of $\mathbf{P}(c-s,s)$ for all c and s, and thus we have the identity $\lambda(c-s,s) = \lambda(c,0)$. We will prove the result when $\forall a \in \mathcal{A} : x_a = 1$. This does not implies that $\mathbf{P}_1 = \mathbf{P}_2$ since the above identity only applies to nonzero coefficients in both matrices. In fact we only have $\mathbf{P}(-1,0) = \mathbf{P}(0,-1)$. We also notice that $\mathbf{P}(s,1-s)$ is identical for all s. We leave as an exercise the case where the x_a are not identical. We notice that $\mathcal{K}$ consists of the tuple $(s,1-s)$ where s is real. We know that

$$C(z_1,z_2) = (1-e^{-z_1})(1-e^{-z_2}) + \sum_{a\in\mathcal{A}} C_a(\pi_1(a)z_1, \pi_2(a)z_2). \tag{104}$$

The issue here is that the $\pi_1(a)$ are not necessarily equal to the $\pi_2(a)$ because they also depend on the other coefficients of matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ which are not tied up by the conjugation property (because their *alter ego* coefficients in the other matrix are null). This implies that we have to consider a new matrix $\mathbf{G}(z)$ whose coefficients $g_{a,b}(z)$ for $(a,b) \in \mathcal{A}^2$ are

$$g_{a,b}(z) = C_a(\pi_1(b)z, \pi_2(b)z).$$

Let $\mathbf{P} = \mathbf{P}(-1,0) = \mathbf{P}(0,-1)$, *i.e.* the matrix whose coefficients are those $\mathbf{P}_1(a|b)$ when $P_1(a|b)P_2(a|b) > 0$, and zero otherwise.

From the functional equation

$$C_a(z_1,z_2) = (1-e^{-z_1})(1-e^{-z_2}) + \sum_{c\in\mathcal{A}} C_c(P(a|c)z_1, P(a|c)z_2).$$

We have the identity

$$g_{a,b}(z) = (1-e^{-\pi_1(b)z})(1-e^{-\pi_2(b)z}) + \sum_{c\in\mathcal{A}} g_{c,b}(P(a|c)z). \tag{105}$$

We have

$$\mathrm{trace}(\mathbf{G}(z)) = \sum_{a\in\mathcal{A}} C_a(\pi_1(a)z, \pi_2(a)z),$$

and then we can rewrite equation (104) for $z_1 = z_2 = z$ as

$$C(z,z) = (1-e^{-z})^2 + \mathrm{trace}(\mathbf{G}(z)).$$

We then compute the Mellin transform $\mathbf{G}^*(s)$ of $\mathbf{G}(z)$, which is a matrix of elements $g^*_{a,b}(s)$ (i.e., Mellin transforms of $g_{a,b}(z)$) that are equal to

$$g^*_{a,b}(s) = r_a(s)\Gamma(s) + \sum_{c\in\mathcal{A}} g^*_{c,b}(s)P(a|c)^{-s}.$$

Here

$$r_a(s) = -\pi_1(a)^{-s} - \pi_2(a)^{-s} + (\pi_1(a) + \pi_2(a))^{-s}$$

is the Mellin transform of the term $(1 - e^{-\pi_1(b)z})(1 - e^{-\pi_2(b)z})$ in (105). Equivalently

$$\mathbf{G}^*(s) = \Gamma(s)\mathbf{r}(s) \otimes \mathbf{1} + \mathbf{P}(s,0)\mathbf{G}^*(s) \tag{106}$$

or $\mathbf{G}(s) = (\mathbf{I} - P(s,0))^{-1}\Gamma(s)\mathbf{r}(s) \otimes \mathbf{1}$ where $\mathbf{r}(s)$ is the vector made of the $r_a(s)$'s. The Mellin transform of $C(z,z)$ which we denote as $c^*(s)$ satisfies

$$c^*(s) = (2^{-s} - 2)\Gamma(s) + \mathrm{trace}(\mathbf{G}^*(s)).$$

The first singularity of $\mathbf{G}(s)$ is the pole of $(\mathbf{I} - \mathbf{P}(s,0))^{-1}$ which is at $s = -\kappa$ such that $\lambda(s,0) = 1$. The only singular term of $\mathbf{G}^*(s)$ at $s = -\kappa$ is on its main eigenvectors:

$$\frac{\Gamma(s)}{1 - \lambda(s,0)} \left(\boldsymbol{\zeta}(s,0) \otimes \mathbf{u}(s,0)\right) \left(\mathbf{r}(s) \otimes \mathbf{1}\right)$$

whose trace is

$$\frac{\Gamma(s)}{1 - \lambda(s,0)} \langle \mathbf{r}(s) | \mathbf{u}(s,0) \rangle \langle \boldsymbol{\zeta}(s,0) | \mathbf{1} \rangle.$$

Thus the residue of trace$(\mathbf{G}^*(s))$ is

$$\gamma_0(-\kappa) = \frac{1}{\lambda'(-\kappa,0)} \langle \mathbf{r}(-\kappa) | \mathbf{u}(-\kappa,0) \rangle \langle \boldsymbol{\zeta}(-\kappa,0) | \mathbf{1} \rangle.$$

The inverse Mellin gives

$$C(z,z) = (1 - e^{-z})^2 + z^\kappa \gamma_0(-\kappa) \left(1 + o(1)\right). \tag{107}$$

When $\mathbf{P}(-1,0)$ is logarithmically rationally related, there are several poles of $(\mathbf{I} - \mathbf{P}(s,0))$ regularly spaced on the vertical axis $\Re(s) = -\kappa$ giving a periodic contribution $z^\kappa Q_0(\log z)$. □

Proof of theorem 7. Let $H_1(s_1,s_2) = \frac{\partial}{\partial s_1}\lambda(s_1,s_2)$ and

$$f(s_1,s_2) = \langle \boldsymbol{\pi}(s_1,s_2) | \mathbf{u}(s_1,s_2) \rangle, \quad g(s_1,s_2) = \langle \boldsymbol{\zeta}(s_1,s_2) | \mathbf{1} \rangle.$$

We first notice that Lemma 11 about the convexity of $L_1(s)$ is still valid since it depends only on general properties of the set $\mathcal{K}$. Define now $L_{j,k}(s)$ implicitly as

$$\lambda_j(-L_{j,k}(s), s) = 1.$$

where $\lambda_j(s_1,s_2)$ is the jth eigenvalues of matrix $\mathbf{P}(s_1,s_2)$, listed in decreasing modulus. The index k indicates that these functions can be polymorphic since the root of the equation $\lambda_j(s_1,s_2) = 1$ for s_2 fixed can be multiple as we have seen in the case when one source is uniform memoryless. For j fixed, each of the functions $L_{j,k}(s)$ are homeomorphic as long as $\lambda_j(L_{j,k}(s),s)$ is non ambiguous *i.e.* the jth eigenvalue has not the same modulus as the previous or next eigenvalues. This would happen only on a discrete set of values s.

Let now for $1 \leq j \leq |\mathcal{A}|$ and $k \in \mathbb{Z}$

$$f_{j,k}(s) = \langle \boldsymbol{\pi}(-L_{j,k}(s), s) | \mathbf{u}_j(-L_{j,k}(s), s) \rangle, \tag{108}$$
$$g_{j,k}(s) = \langle \boldsymbol{\zeta}_j(-L_{j,k}(s), s) | \mathbf{1} \rangle. \tag{109}$$

Then

$$I(z,\rho) = \tag{110}$$
$$\frac{1}{2i\pi} \int_{\Re(s)=\rho} \sum_{k\in\mathbb{Z}} \sum_{j=1}^{|\mathcal{A}|} \frac{f_{j,k}(s) g_{j,k}(s) \Gamma(-L_{j,k}(s)) \Gamma(s)}{-\frac{\partial}{\partial s_1}\lambda_j(-L_{j,k}(s), s)} z^{L_{j,k}(s)-s} ds$$

and

$$J(z,\rho)=\frac{1}{2i\pi}\int_{\Re(s)=c_2}\widehat{C}(0,s)\Gamma(s)z^{-s}ds. \tag{111}$$

With these new definitions the expression (79) in Lemma 13 is still valid, that is for all $M>0$:

$$C(z,z)=1+I(z,\rho)+J(z,\rho)+O(z^{-M}).$$

The proof is indeed the same, the pole cancellations occur the same way and the identity between residues is formally the same.

We know that for all k we have $\lambda_1(-L_{1,k}(s),s)=1$. We assume that $k=0$ defines the branch where $L_{1,0}(s)$ is real when s is real. To simplify we denote $L(s)=L_{1,0}(s)$. We then move the integration path to the value $s=c_2$ which attains the minimum of $L(s)-s$ at $s=\kappa$. We know that the

$$J(z,c_2)=O(z^{\kappa-\epsilon})$$

for $\epsilon>0$ such that $\kappa-\epsilon>L(0)$. Therefore

$$C(z,z)=1+I(z,c_2)+O(z^{\kappa-\epsilon})$$

as in the case with uniform one source. Since $z^{L(s)-s}$ is at a minimum for real values of s, we have again a saddle point for $\int_{\Re(s)=c_2}\mu(s)z^{L(s)-s}ds$.

Thus we arrive to two cases: (i) either $c_1>0$ or $c_2>0$ or (ii) c_1 and c_2 are both negative. In the first case the condition of Theorem 7 applies. In the second case the condition of Theorem 8 applies. We consider the case $c_2>0$ (case $c_1>0$ is symmetric). Moving the integration toward $\Re(s)=c_2$ one meets the pole of function $\Gamma(s)$ at zero.

When we meet the pole at $s=0$ by moving ρ toward the positive value we obtain a residue from $I(z,\rho)$ equal to $\sum_j H_j(0,z)$ and a residue from $J(z,\rho)$ equal to $\widehat{C}(0,0)$. Notice that $H_1(0,z)=\Omega(z^{c_0})$ since $c_0=L(0)>0$ while the residue from $J(z,\rho)$ is negligible. The function $H_1(0,z)$ turns out to be the leading term since the other terms are of order $z^{L_j(0)}$ for $j>1$ and $\Re(L_j(0))<L_1(0)$. By moving again the integration path with $\Re(s)>0$ we arrive at $\Re(s)=c_2$ thus

$$C(z,z)=\sum_j H_j(0,z)+I(z,c_2)+O(z^{-M}).$$

We know from the previous discussion that $I(z,c_2)=O(z^{\kappa})$ which is of order smaller than $z^{L(0)}$ per definition of κ. Therefore we have $C(z,z)=1+H_1(0,z)+O(z^{c_0-\epsilon})$ for some $\epsilon>0$. Notice that

$$H_1(0,z)=\frac{f_1(0)g_1(0)}{\lambda_1(0)}z^{c_0}+Q_1(\log z)z^{c_0}$$

where $Q_1(.)$ is a periodic function of periodic $\log|\mathcal{A}|$ and of mean 0 with small amplitude.

Recapitulating all cases of Theorem 7:

(i) when $\mathbf{P}(-1,0)$ is not logarithmically related, then

$$C_{n,n}=\gamma_1(c_0,0)n^{-c_0}(1+o(1))$$

with

$$\gamma_1(s_1,s_2)=\frac{f(s_1,s_2)g(s_1,s_2)(1+s_1)\Gamma(s_1)}{H_1(s_1,s_2)}.$$

(ii) When $\mathbf{P}(-1,0)$ is logarithmically commensurable

$$C_{n,n}=n^{-c_0}\sum_{k\in\mathbb{Z}}n^{2ik\pi\nu}\gamma(c_0+2ik\pi\nu,0)+O(n^{c_0-\epsilon})$$

where ν is the root of $\mathbf{P}_1$. □

Remark Remember that when all coefficients of $\mathbf{P}_2$ are non-negative $\mathbf{P}(-1,0) = \mathbf{P}_1$ which is not necessarily the case when $\mathbf{P}_2$ has some null coefficients.

Proof of Theorems 8 and 9. We need the following lemma which is basically equivalent of Lemmas 17 and 18 developed in the special case.

Lemma 21. *There exists $\epsilon > 0$ such that for all integers $j > 1$ and for all integers k the quantities*

$$\int_{\Re(s)=c_2} \frac{f_{j,k}(s)g_{j,k}(s)\Gamma(-L_{j,k}(s))\Gamma(s)}{-\frac{\partial}{\partial s_1}\lambda_j(-L_{j,k}(s),s)} z^{L_{j,k}(s)-s} ds \tag{112}$$

are uniformly $O(z^{\kappa-\epsilon})$.

Proof. The proof consists of showing that there exist $\epsilon > 0$ such that if $\lambda_j(s_1,s_2) = 1$ with $\Re(s_2) = c_2$ then $\Re(s_1) > c_1 + \epsilon$. First we prove that $\Re(s_1) \geq c_1$. We know that $|\lambda_1(s_1,s_2)| \geq |\lambda_j(s_1,s_2)| = 1$. Since $|\lambda_1(s_1,s_2)| \leq \lambda_1(\Re(s_1),\Re(s_2))$. We have $\Re(s_2) = c_2$, thus the inequality $\lambda_1(\Re(s_1),c_2) \geq 1$ implies that $\Re(s_1) \geq c_1$ since $\lambda_1(\Re(s_1),\Re(s_2))$ is strictly increasing in $\Re(s_1)$ and $\Re(s_2)$.

Second we prove the existence of ϵ. By absurdum we assume that there is a sequence of complex numbers (x_k,y_k) such that $\lambda_j(x_k,y_k) = 1$ and $\Re(y_k) = c_2$ and $\Re(x_k) \to c_1$ with $\Re(x_k) \geq c_1$. We know that $|\lambda_1(x_k,y_k)| \geq |\lambda_j(x_k,y_k)| = 1$. From the inequality

$$\lambda_1(\Re(x_k),\Re(y_k)) \geq |\lambda_1(x_k,y_k)| \geq 1$$

we get that $|\lambda(x_k,y_k)| \to 1$, since $(\Re(x_k),\Re(y_k)) \to (c_1,c_2)$. This would imply that

$$\left|\frac{\lambda_j(x_k,y_k)}{\lambda_1(x_k,y_k)}\right| \to 1,$$

which contradicts Lemma 10. □

The above lemma fills the gap necessary to establish of Theorems 8 and 9 by following the footsteps of the proofs of Theorems 4 and 5. □

References

[1] G. Burnside, D. Milioris and P. Jacquet, One Day in Twitter: Topic Detection Via Joint Complexity, *Proceedings of SNOW 2014 Data Challenge, WWW'14, Seoul, South Korea*, April 2014.

[2] P. Flajolet, X. Gourdon, and P. Dumas, Mellin Transforms and Asymptotics: Harmonic sums, *Theoretical Computer Science*, 144, 3-58, 1995.

[3] V. Becher and P. A. Heiber, A linearly computable measure of string complexity, *Theoretical Computer Science*, 438, 62-73, 2012.

[4] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2008.

[5] Ilie, L., Yu, S., and Zhang, K. Repetition Complexity of Words In *Proc. COCOON* 320-329, 2002.

[6] P. Jacquet, Common words between two random strings, *IEEE Intl. Symposium on Information Theory*, 1495-1499, 2007.

[7] P. Jacquet, and W. Szpankowski, Autocorrelation on Words and Its Applications. Analysis of Suffix Trees by String-Ruler Approach, *J. Combinatorial Theory Ser. A*, 66, 237-269, 1994.

[8] P. Jacquet, and W. Szpankowski, Analytical DePoissonization and Its Applications, *Theoretical Computer Science*, 201, 1-62, 1998.

[9] P. Jacquet and W. Szpankowski, Joint String Complexity for Markov Sources, *23rd International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, AofA'12, *DMTCS Proc.*, 303-322, Montreal, 2012.

[10] P. Jacquet, and W. Szpankowski, *Analytic Pattern Matching: From DNA to Twitter*, Cambridge University Press, Cambridge, 2015.

[11] P. Jacquet, W. Szpankowski, and J. Tang, Average Profile of the Lempel-Ziv Parsing Scheme for a Markovian Source, *Algorithmica*, 31, 318-360, 2001.

[12] P. Jacquet and W. Szpankowski, Average Size of a Suffix Tree for Markov Sources, *27th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, AofA'16, Krakow, 2016.

[13] S. Janson, S. Lonardi and W. Szpankowski, On Average Sequence Complexity, *Theoretical Computer Science*, 326, 213-227, 2004.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[15] K. Leckey, R. Neininger, and W. Szpankowski, Towards More Realistic Probabilistic Models for Data Structures: The External Path Length in Tries under the Markov Model, *SIAM-ACM Symposium on Discrete Algorithms* (SODA 2013), 877-886, New Orleans, 2013.

[16] Li, M., and Vitanyi, P. *Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, Berlin, Aug. 1993.

[17] N. Merhav and W. Szpankowski, Average Redundancy of the Shannon Code for Markov Sources, *IEEE Trans. Information Theory*, 59, 7186-7193, 2013.

[18] D. Milioris and P. Jacquet, Joint Sequence Complexity Analysis: Application to Social Networks Information Flow, *Bell Labs Technical Journal*, Vol. 18, No. 4, 2014

[19] Niederreiter, H., Some computable complexity measures for binary sequences, In *Sequences and Their Applications*, Eds. C. Ding, T. Hellseth and H. Niederreiter Springer Verlag, 67-78, 1999.

[20] J. Fayolle, M. Ward, Analysis of the average depth in a suffix tree under a Markov model DMTCS Proceedings of AofA 2005.

[21] B. Pittel, Asymptotic Growth of a Class of Random Trees, *Annals of Probability*, 18, 414-427, 1985.

[22] M. Régnier and W. Szpankowski, On pattern frequency occurrences in a Markovian sequence, *Algorithmica*, 22, 631-649, 1998.

[23] W. Szpankowski, *Analysis of Algorithms on Sequences*, John Wiley, New York, 2001.

[24] J. Ziv, On classification with empirically observed statistics and universal data compression, *IEEE Trans. Information Theory*, 34, 278-286, 1988.

Appendix

Proof of Lemma 2. Let $\forall(a,b) \in \mathcal{A}^2$ $P_1(a|b) > 0$. Thus $(0,-1) \in \mathcal{K}$, but since some $P_2(a|b)$ may be zero, the point $(-1,0)$ may not be $\mathcal{K}$. But if $(-1,s_2) \in \mathcal{K}$ with $s_2 \in \mathbb{R}$ then $s_2 \geq 0$, otherwise since $\mathbf{P}(-1,s_2) \leq \mathbf{P}_1$ coefficientwise, then $\lambda(-1,s_2) < 1$. Similarly if $(s_1,0) \in \mathcal{K}$ then $s_1 \geq -1$.

We know that the curve $(-s_1,-s_2)$ is convex for $(s_1,s_2) \in \mathcal{K}$, so is the curve $(s_1,-s_1-s_2)$, then $-s_1-s_2$ is a function of s_1, say $a(s_1)$. We have $a(0) = 1$ and $a(-1) \leq 1$. Thus the minimum value of $a(s_1)$ which is κ is attained on c_1 which must satisfies $c_1 \leq 0$ We then have $c_1 < 0$ when the curve is strictly convex.

Similarly, the curve $(s_2,-s_1-s_2)$ is convex, so is the function $b(s_2) = -s_1-s_2$. Since $b(-1) = 1$ and $b(0) \leq 1$ the minimum is necessarily attained on $c_2 \geq -1$, and $c_2 > -1$ when it is strictly convex. $\square$

Proof of Lemma 4. In order to prove Lemma 4 we adopt here the following general double depoissonization lemma that is proved in [10] (see Lemmma 10.3.4 in Chapter 10).

Lemma 22. *Let $a_{n,m}$ be a two-dimensional (double) sequence of complex numbers. We define the double Poisson transform $f(z_1, z_2)$ of $a_{n,m}$ as*

$$f(z_1, z_2) = \sum_{n,m\geq 0} a_{n,m} \frac{z_1^n}{n!} \frac{z_1^m}{m!} e^{-z_1-z_2}.$$

Let now $\mathcal{S}_\theta$ be a cone of angle θ around the real axis. Assume that there exist $B > 0$, $D > 0$, $\alpha < 1$ and β such that for $|z_1|$, $|z_2| \to \infty$:

(i) *if $z_1, z_2 \in \mathcal{S}_\theta$ then $|f(z_1, z_2)| = B(|z_1|^\beta + |z_2|^\beta)$;*

(ii) *if $z_1, z_2 \notin \mathcal{S}_\theta$ then $|f(z_1, z_2)e^{z_1+z_2}| = De^{\alpha|z_1|+\alpha|z_2|}$;*

(iii) *if $z_i \in \mathcal{S}_\theta$ and $z_j \notin \mathcal{S}_\theta$ for $\{i, j\} = \{1, 2\}$ and $|f(z_1, z_2)e^{z_j}| < D|z_i|^\beta e^{\alpha|z_j|}$.*

Then

$$a_{n,m} = f(n, m) + O\left(\frac{n^\beta}{m} + \frac{m^\beta}{n}\right)$$

for large m and n.

Just to prove the Lemma 4, we need to establish three conditions (i)-(iii) of Lemma 22. We accomplish it through a generalization of the so called *increasing domain* approach discussed in [8, 23].

We first prove the lemma for the generating functions $C_a(z_1, z_2)$ for every $a \in \mathcal{A}$. Assume now that $\rho = \max_{(a,b)\in\mathcal{A}^2,i\in\{1,2\}}\{P_i(a|b)\}$. We denote by $\mathcal{S}_k$ part of the cone $\mathcal{S}_\theta$ that contains points such that $|z| < \rho^{-k}$. Notice that $\mathcal{S}_k \subset \mathcal{S}_{k+1}$ for all integer k. We also notice $C(z_1, z_2) = O((|z_1| + |z_2|)^2)$ when $z_1, z_2 \to 0$, therefore we can define

$$B_k = \max_{a\in\mathcal{A},(z_1,z_2)\in\mathcal{S}_k\times\mathcal{S}_k} \frac{|C_a(z_1, z_2)|}{|z_1| + |z_2|} < \infty \ .$$

We use the functional equation

$$C_b(z_1, z_2) = (1 - (1 + z_1)e^{-z_1})(1 - (1 + z_2)e^{-z_2}) + \sum_{a\in\mathcal{A}} C_a\left(P_1(a|b)z_1, P_2(a|b)z_2\right). \tag{113}$$

In the above equation, we notice that if $(z_1, z_2) \in \mathcal{S}_{k+1} \times \mathcal{S}_{k+1} - \mathcal{S}_k \times \mathcal{S}_k$, then for all $(a, b) \in \mathcal{A}^2$ $(P_1(a|b)z_1, P_2(a|b)z_2)$ are in $\mathcal{S}_k \times \mathcal{S}_k$ and therefore we have for some fixed $\beta > 0$ and for all $b \in \mathcal{A}$:

$$|C_b(z_1, z_2)| \leq B_k(\sum_{a\in\mathcal{A}} P_1(a|b)|z_1| + P_2(a|b)|z_2|) + \beta = B_k(|z_1| + |z_2|) + \beta \tag{114}$$

since $|1 - (1 + z_i)e^{-z_i}|$ is uniformly bounded for all integers k by some $\sqrt{\beta}$ for both $i \in \{1, 2\}$ when $(z_1, z_2) \in \mathcal{S}_k$. Thus, we can derive the following recurrent inequality:

$$B_{k+1} \leq B_k + \beta \max_{(z_1,z_2)\in\mathcal{S}_{k+1}\times\mathcal{S}_{k+1}-\mathcal{S}_k\times\mathcal{S}_k} \{\frac{1}{|z_1| + |z_2|}\} = B_k + \beta\rho^k \ . \tag{115}$$

We should notice that

$$\min_{(z_1,z_2)\in\mathcal{S}_{k+1}\times\mathcal{S}_{k+1}-\mathcal{S}_k\times\mathcal{S}_k} \{|z_1| + |z_2|\} = \rho^{-k} \tag{116}$$

because one of the number z_i has modulus greater than ρ^{-k}. It turns out that $\lim_{k\to\infty} B_k < \infty$, establishing condition (i) of the double depoissonization Lemma 22.

Now we are going to establish condition (iii). To this end we define $\mathcal{G}$ as the complementary cone of $\mathcal{S}_\theta$ and $\mathcal{G}_k$ as the portion made of the point of modulus smaller than ρ^{-k}. We will use $\cos\theta < \alpha < 1$, therefore $\forall z \in \mathcal{G}$: $|e^z| < e^{\alpha|z|}$. We define D_k as

$$D_k = \max_{a\in\mathcal{A},(z_1,z_2)\in\mathcal{G}_k\times\mathcal{G}_k} \frac{|C_a(z_1, z_2)e^{z_1+z_2}|}{\exp(\alpha|z_1| + \alpha|z_2|)} \ . \tag{117}$$

We define $G_a(z_1,z_2)=C_a(z_1,z_2)e^{z_1+z_2}$, we have the following equation

$$G_b(z_1,z_2)=(e^{z_1}-1-z_1)(e^{z_2}-1-z_2)+\sum_{a\in\mathcal{A}}C_a\left(P_1(a|b)z_1,P_2(a|b)z_2\right)e^{1-P_1(a|b)z_1+(1-P_2(a|b))z_2}\ . \quad (118)$$

We notice that if $(z_1,z_2)\in\mathcal{G}_{k+1}\times\mathcal{G}_{k+1}-\mathcal{G}_k\times\mathcal{G}_k$, then all $(P_1(a|b)z_1,P_2(a|b)z_2)$ are in $\mathcal{G}_k\times\mathcal{G}_k$ and therefore we have for all $b\in\mathcal{A}$:

$$\begin{aligned}|G_b(z_1,z_2)| \quad &\leq \quad D_k\left(\sum_{a\in\mathcal{A}}\exp\left((P_1(a|b)\alpha+(1-P_1(a|b))\cos\theta)|z_1|+(P_2(a|b)\alpha+(1-P_2(a|b))\cos\theta)|z_2|\right)\right)\\ &\quad +(e^{\cos\theta|z_1|}+1+|z_1|)(e^{\cos\theta|z_2|}+1+|z_2|).\end{aligned}$$

We notice that $\forall(a,b)\in\mathcal{A}^2$ and $\forall i\in\{1,2\}$:

$$P_i(a|b)\alpha+(1-P_i(a|b))\cos\theta-\alpha\leq-(1-\rho)(\alpha-\cos\theta)\ , \quad (119)$$

We also have $e^{\cos\theta|z_i|}+1+|z_i|\leq e^{\cos\theta|z_i|}(2+\frac{1}{e\cos\theta})$, therefore

$$\frac{|G_b(z_1,z_2)|}{\exp(\alpha(|z_1|+|z_2|))}\leq D_k|\mathcal{A}|e^{-(1-\rho)(\alpha-\cos\theta)(|z_1|+|z_2|)}+(2+\frac{1}{e\cos\theta})^2e^{-(\alpha-\cos\theta)(|z_1|+|z_2|)}\ . \quad (120)$$

Since $(z_1,z_2)\in\mathcal{G}_{k+1}\times\mathcal{G}_{k+1}-\mathcal{G}_k\times\mathcal{G}_k$ implies $|z_1|+|z_2|\geq\rho^{-k}$ it follows

$$D_{k+1}\leq\max\left\{D_k,|\mathcal{A}|D_ke^{-(1-\rho)(\alpha-\cos\theta)\rho^{-k}}+(2+\frac{1}{e\cos\theta})^2e^{-(\alpha-\cos\theta)\rho^{-k}}\right\}\ . \quad (121)$$

We clearly have $\lim_{k\to\infty}D_k<\infty$ and condition (iii) is established.

The proof of condition (ii) for z_1 and z_2 being in $\mathcal{S}_\theta$ and $\mathcal{G}$ is a mixture of the above proofs. Furthermore, the proof about the unconditional generating function $C(z_1,z_2)$ is a trivial extension. □

Proof of Lemma 5. Let $\mathbf{u}=(u_a)_{a\in\mathcal{A}}$ be the right eigenvector of $\mathbf{M}$ and $(v_a)_{a\in\mathcal{A}}$ be the right eigenvector of $\mathbf{Q}$. Let also $v_a=x_au_a$. If 1 is the eigenvalue, we have for all $c\in\mathcal{A}$:

$$(1-e^{i\theta_{cc}}m_{cc})u_c=\sum_{b\neq c}m_{cb}u_be^{i\theta_{cb}}\frac{x_b}{x_c}\ . \quad (122)$$

If $e^{i\theta_{cc}}\neq1$, then

$$|(1-e^{i\theta_{cc}}m_{cc})u_c|>(1-m_{cc})u_c. \quad (123)$$

By the Perron-Frobenius theorem all u_a are real non negative. Suppose that $|x_c|=\max_{a\in\mathcal{A}}\{|x_a|\}$. If $\exists d\in\mathcal{A}$: $\frac{|x_d|}{|x_c|}<1$ or if $(b,b')\in(\mathcal{A}-\{c\})^2$: $e^{i\theta_{cb}}\frac{x_b}{x_c}\neq e^{i\theta_{cb'}}\frac{x'_b}{x_c}$. Then

$$\left|\sum_{b\neq c}m_{cb}u_be^{i\theta_{cb}}\frac{x_b}{x_c}\right|<\sum_{b\neq c}m_{cb}u_b\ . \quad (124)$$

But we also know that

$$(1-m_{cc})u_c=\sum_{b\neq c}m_{cb}u_b\ . \quad (125)$$

Therefore, we have $e^{i\theta_{cc}}=1$ and for all $b\in\mathcal{A}$: $|x_b|=|x_c|$, and for all $(b,b')\in(\mathcal{A}-\{c\})^2$: $e^{i\theta_{cb}}\frac{x_b}{x_c}=e^{i\theta_{cb'}}\frac{x'_b}{x_c}$. But since for all $b\in\mathcal{A}$ $|x_b|=|x_c|$ every symbol in $\mathcal{A}$ can play the role of c. Since for all $c\in\mathcal{A}$

$$(1-m_{cc})=\sum_{b\neq c}m_{cb}u_be^{i\theta_{cb}}\frac{x_b}{x_c}=\sum_{b\neq c}m_{cb}u_b\ , \quad (126)$$

we simply have $\forall(a,b)\in\mathcal{A}$: $e^{i\theta_{ab}}\frac{x_b}{x_a}=1$. Denoting $x_a=e^{i\theta_a}$ we prove the expected result. The converse proposition is immediate. □

Proof of Lemmas 1 and 7. We call $\tilde{\mathcal{K}}$ the set of real tuples such that $\lambda(s_1, s_2) \leq 1$. The set $\mathcal{K}$ is the topological border of $\tilde{\mathcal{K}}$ and since $\lambda(s_1, s_2)$ decreases when s_1 or s_2 decrease, it is the upper border. We will show that $\tilde{\mathcal{K}}$ is a convex set and thus its upper border is concave. Let (x_1, x_2) and (y_1, y_2) be two elements of $\tilde{\mathcal{K}}$ and α and β two non negative real numbers such that $\alpha + \beta = 1$. We want to prove that $(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2) \in \tilde{\mathcal{K}}$.

By construction

$$\mathbf{P}(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2) = \mathbf{P}(\alpha x_1, \alpha x_2) \star \mathbf{P}(\beta y_1, \beta y_2)$$

where $\star$ denotes the Schur product. For $(s_1, s_2) \in \tilde{\mathcal{K}}$ let $\mathbf{u}(s_1, s_2)$ the right main eigenvector of $\mathbf{P}(s_1, s_2)$, *i.e.*

$$\mathbf{P}(s_1, s_2)\mathbf{u}(s_1, s_2) = \lambda(s_1, s_2)\mathbf{u}(s_1, s_2).$$

We know that $\lambda(s_1, s_2) \leq 1$ therefore

$$\mathbf{P}(s_1, s_2)\mathbf{u}(s_1, s_2) \leq \mathbf{u}(s_1, s_2)$$

coefficientwise. Let $\mathbf{u}(s_1, s_2)^{\star\alpha}$ denotes the vector $\mathbf{u}(s_1, s_2)$ with all its coefficients raised to power α. We want to give an estimate of

$$\mathbf{P}(\alpha x_1, \alpha x_2) \star \mathbf{P}(\beta y_1, \beta y_2)$$

applied to the vector

$$\mathbf{u}(x_1, x_2)^{\star\alpha} \star \mathbf{u}(y_1, y_2)^{\star\beta}.$$

Let $a \in \mathcal{A}$ the coefficient of the vector

$$\mathbf{P}(\alpha x_1, \alpha x_2) \star \mathbf{P}(\beta y_1, \beta y_2)\mathbf{u}(x_1, x_2)^{\star\alpha} \star \mathbf{u}(y_1, y_2)^{\star\beta}$$

corresponding to symbol a is equal to

$$\sum_{b\in\mathcal{A}} u_b(x_1, x_2)^\alpha P_1(a|b)^{-\alpha x_1} P_2(a|b)^{-\alpha x_2} u_b(y_1, y_2)^\beta P_1(a|b)^{-\beta y_1}$$

$$P_2(a|b)^{-\beta y_2} \sum_{b\in\mathcal{A}} u_b(x_1, x_2)^\alpha P_1(a|b)^{-\alpha x_1} P_2(a|b)^{-\alpha x_2} u_b(y_1, y_2)^\beta P_1(a|b)^{-\beta y_1} P_2(a|b)^{-\beta y_2}.$$

Using Hölder inequality, the above quantity is smaller than

$$\left(\sum_{b\in\mathcal{A}} u_b(x_1, x_2) P_1(a|b)^{-x_1} P_2(a|b)^{-x_2}\right)^\alpha \left(\sum_{b'\in\mathcal{A}} u_b(y_1, y_2) P_1(a|b')^{-y_1} P_2(a|b')^{-y_2}\right)^\beta. \tag{127}$$

The above terms are respectively $\lambda(x_1, x_2)u_a(x_1, x_2)$ and $\lambda(y_1, y_2)u_a(y_1, y_2)$. Therefore the vector

$$\mathbf{P}(\alpha x_1, \alpha x_2) \star \mathbf{P}(\beta y_1, \beta y_2)\mathbf{u}(x_1, x_2)^{\star\alpha} \star \mathbf{u}(y_1, y_2)^{\star\beta}$$

is coefficientwise smaller than

$$\lambda^\alpha(x_1, x_2)\lambda^\beta(y_1, y_2)\mathbf{u}(x_1, x_2)^{\star\alpha} \star \mathbf{u}(y_1, y_2)^{\star\beta}.$$

Since $\lambda^\alpha(x_1, x_2)\lambda(^\beta(y_1, y_2) \leq 1$ by Perron-Frobenius the main eigenvalue of $\mathbf{P}(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2)$ is smaller than or equal to 1, consequently $(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2) \in \bar{\mathcal{K}}$.

The Hölder inequality is an equality if and only if the vectors $(u_a(x_1, x_2)\mathbf{P}(x_1, x_2))_{a\in\mathcal{A}}$ and $(u_a(y_1, y_2)\mathbf{P}(y_1, y_2))_{a\in\mathcal{A}}$ are colinear, which happens when $\mathbf{P}(x_1, x_2)$ and $\mathbf{P}(y_1, y_2)$ are conjugate, which is equivalent to the fact that $\mathbf{P}_1$ and $\mathbf{P}_2$ are conjugate (on the coefficients which are non zero). $\square$

Proof of Lemma 10. Consider the matrix $\frac{1}{\lambda(x_k,y_k)}\mathbf{P}(x_k, y_k)$. Since the coefficients of this matrix are bounded, there is no loss in generality to consider the sequence of matrices converging to a matrix $\mathbf{M}$. The matrix $\mathbf{M}$ and matrix $\mathbf{Q} = \mathbf{P}(c_1, c_2)$, as defined in Lemma 5, are imaginary conjugate *i.e.* the coefficients of $\mathbf{M}$ are of the form

$$e^{i(\theta_a-\theta_b)}P_1(a|b)^{-c_1}P_2(a|b)^{-c_2} \tag{128}$$

for some vector of real numbers θ_a. Therefore, $\mathbf{M}$ and $\mathbf{P}(c_1, c_2)$ have the same spectrum. The spectrum of $\frac{1}{\lambda(x_k,y_k)}\mathbf{P}(x_k, y_k)$ converges to the spectrum of $\mathbf{M}$. Furthermore, the right eigenvector $\mathbf{u}(x_k, y_k)$ converges to the vector $e^{i\theta_a}u_a(c_1, c_2)$ and the left eigenvector $\boldsymbol{\zeta}(x_k, y_k)$ converges to $e^{-i\theta_a}\zeta_a(c_1, c_2)$.

For any pair of complex numbers (s_1, s_2) we have the identity

$$\frac{1}{\lambda(x_k, y_k)}\mathbf{P}(x_k + s_1, y_k + s_2) = \frac{1}{\lambda(x_k, y_k)}\mathbf{P}(x_k, y_k) * \mathbf{P}(s_1, s_2)\ . \tag{129}$$

Thus $\frac{1}{\lambda(x_k,y_k)}\mathbf{P}(x_k+s_1, y_k+s_2)$ converges to $\mathbf{M}*\mathbf{P}(s_1, s_2)$ and is conjugate to $\mathbf{P}(c_1+s_1, c_2+s_2)$. Since the eigen spectrum of $\frac{1}{\lambda(x_k,y_k)}\mathbf{P}(x_k + s_1, y_k + s_2)$ converges to the eigen spectrum of $\mathbf{P}(c_1 + s_1, c_2 + s_2)$, thus we have $\lambda(x_k+s_1, y_k+s_2) \to \lambda(c_1+s_1, c_2+s_2)$. We also have $|\lambda(x_k+s_1, y_k+s_2)| > |\lambda_2(x_k+s_1, y_k+s_2)|$ when k is large enough with (s_1, s_2) in the complex neighborhood $\mathcal{U}^2$ which implies the analyticity of $\lambda(x_k + s_1, y_k + s_2)$. Thus by Ascoli theorem the derivatives converge, too. $\square$